



Visualización de Datos de la Matriz de Producción Eléctrica Costarricense

Trabajo Final de Graduación

Tecnológico de Costa Rica

Escuela de Ingeniería en Diseño Industrial

**Visualización de Datos de la Matriz de Producción Eléctrica
Costarricense**

Trabajo Final de Graduación para optar por el título de
Ingeniero en Diseño Industrial con énfasis en Comunicación
Visual con grado de Licenciatura

Estudiante: Ing. Josué Porras Fernández

Profesor Asesor Ph.D Franklin Hernández-Castro

Junio 2019

NOTA FINAL

Trabajo Final de Proyecto de Graduación_Licenciatura
Ingeniería en Diseño Industrial_Comunicación Visual

El Trabajo Final de Graduación presentado por el estudiante Josué Porras Fernández titulado:

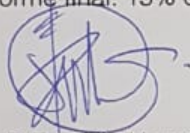
Visualización de Datos de la Matriz de Producción Eléctrica Costarricense

ha sido defendido públicamente el día Lunes 3 de junio del año 2019 ante el Tribunal Evaluador conformado por Máster Ivonne Madrigal y Máster María del Carmen Valverde y su profesor Asesor Ph.D Franklin Hernández.

Han sido evaluados los entregables y su calificación es la siguiente:

Profesor Asesor_30%

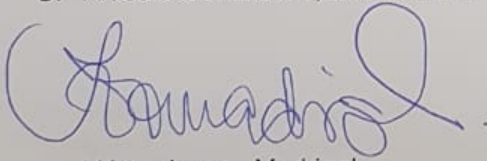
- A. Desempeño estudiante: 15% de 15%
- B. Informe final: 15% de 15%



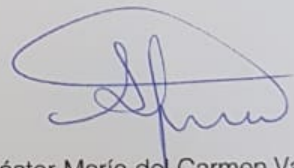
Ph.D Franklin Hernández

Tribunal Evaluador_70%

- A. Preguntas: 8,75% de 10
- B. Presentación final: 40% de 50%
- C. Artículo Científico: 5,25% de 10%

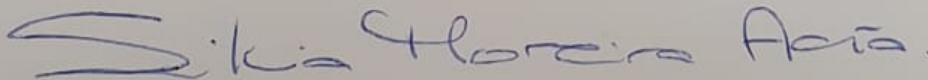


Máster Ivonne Madrigal



Máster María del Carmen Valverde

La sumatoria anterior es de 84 de 100 y la Nota Final del Trabajo Final de Graduación es de: 85.
Dicho Trabajo Final de Graduación ha sido aprobado.



DI. Silvia Elena Moreira Acuña, M.Sc
Coordinadora Trabajo Final de Graduación_Licenciatura

Miércoles 5 de junio 2019, Cartago

Tabla de Contenidos

Tabla de Contenidos	1
Índice de Figuras	4
Índice de Tablas	7
Agradecimientos	8
Dedicatoria	9
1. Introducción	10
2. Antecedentes - Indicadores	11
3. Definición del problema	14
3.1 Planteamiento	14
3.1 Árbol del Problema.....	15
4. Definición de Objetivos	16
Objetivo General	16
Objetivos Específicos	16
5. Preguntas de Investigación	17
Pregunta Objetivo	17
Preguntas Secundarias.....	17
6. Diagnóstico de la situación	18
6.1 Sistema Eléctrico Nacional (SEN)	18
6.1.1 Sistema único e interconectado	18
6.1.2 Parques eléctricos	19
6.1.3 Sistema Mixto	19
6.1.4 Matriz Eléctrica Costarricense	20
6.2 Generación de Energía	21
6.2.1 Planificación de la producción energética ..	22
7. Marco Teórico	24
7.1 La Generación de Energía	24
8.Marco Metodológico	33

8.1 Investigación	33
8.2 Metodología	33
10. Análisis de Referenciales	35
10.1 Referenciales	35
10.1.1 The daily routines of famous creative people	35
10.1.2 What's really warming the world	36
10.1.3 The complete history of the NFL	37
10.1.4 The year in news	38
10.1.5 Terrorist attacks	39
10.2 Conclusiones y mínimos comunes.	39
11. Análisis de los datos	40
11.1 Desarrollo de Hipótesis	41
11.1.1 Visualizar todas las variables a una hora específica	41
11.1.2 Comparar el mismo día de la semana en diferentes meses y horas	42
11.2 Data Mining	44
11.3 Data Cleaning	46
Definición de los Set de Datos.....	46
Dimensión de las variables	48
12. Paradigmas de visualización	50
12.1 Paradigmas Seleccionados	50
12.1.1 Stream Graph	50
12.2 Stacked Area Graph.....	51
12.3 Parallel Coordinates.....	52
12.2 Casos de Estudio.....	53
12.2.1 Visualizar cada central eléctrica.....	53
12.2.2 Subutilización de las centrales	56
12.2.3 Relación entre predespacho y energía real	59
12.2.4 Época del año con mayor producción por cada central	61

12.3 Matriz de Selección del Paradigma	63
12.3.1 Matriz de evaluación según casos de estudio	63
12.3.1 Matriz de evaluación según requerimientos de visualización	64
13. Implementación	66
13.1 Definición de aspectos visuales	66
Cromática	66
Líneas.....	67
13.2 Visualización de Datos	67
14. Validación y resultados.....	72
14.1 Casos de estudio	72
14.1.1 Determinar cual central tiene el grado mas alto de sub utilización.....	72
14.1.2 Determinar cual central produjo mas energía...72	
14.1.3 Buscar una central eléctrica por su nombre72	
14.1.4 Conocer la fecha y hora específica de las coordenadas 72	
14.1.5 Comparar dos centrales en un periodo de tiempo específico.....	73
14.1.6 Comparar dos variables por separado	74
14.1.7 Visualizar el mismo día de la semana a través un mes a una hora específica.....	74
14.1.8 Visualizar el comportamiento de un tipo de fuente en un periodo específico de tiempo.	75
14.1.9 Visualizar en cual época del año se produce mas energía para un tipo de fuente en específico	76
14. Hallazgos	77
15. Conclusiones	78
16. Trabajo a futuro	79
17. Referencias	80

Índice de Figuras

Figura 01. Aumento de la producción energética limpia del año 2014 al 2017. Casa Presidencial 2017 -----	11
Figura 02. Distribución de la generación eléctrica nacional para el año 2017, por tipo de recurso utilizado -----	12
Figura 03. Árbol del Problema -----	15
Figura 04. Sistema eléctrico nacional por central y tipo de fuente utilizada en el año 2018, según datos extraídos del CENCE. -----	20
Figura 05. Porcentaje de participación según tipo de entidad productora. Datos al 31 de diciembre del 2014. -----	21
Figura 06. Capacidad instalada según tipo de recurso utilizado por los diferentes productores. -----	22
Figura 07. Flujo de planificación y producción energética entre el CENSE y productores del mercado. -----	23
Figura 08. Stacked Area Graph. Fuente: The Data Visualisation Catalogue.-----	27
Figura 09. Bubble Chart. Fuente: The Data Visualisation Catalogue -----	27
Figura 10. Heatmap. Fuente: The Data Visualisation Catalogue. -	28
Figura 11. Histograma. Fuente: The Data Visualisation Catalogue, Severino Ribecca. -----	28
Figura 12. Gráfico de líneas. Fuente: The Data Visualisation Catalogue. -----	29
Figura 13. Causas de mortalidad del ejército del este en la guerra de Crimea. Autora: Florence Nightingale. -----	29
Figura 15. Gráfico de Flujo. Fuente: The Data Visualisation Catalogue. -----	30
Figura 16. Parallel Coordinates. Fuente: The Data Visualisation Catalogue, Severino Ribecca. -----	31
Figura 19. Bar Graph. The daily routines of famous creative people. Fuente:Podio -----	35
Figura 20. Line Graph. What's really warming the world?. Fuente: Bloombergs Business Work, 2015.-----	36

Figura 21. Line Graph. The Complete History of the NFL. Fuente: Project Five Thirty Eight. -----	37
Figura 22. Stream Graph. The Year in News. Fuente: Echelon Insights, 2015.-----	38
Figura 21. Bump Chart. Terrorist Attacks. Fuente: Density Design -----	39
Figura 22. Generación real VS predespacho de la planta hidroeléctrica Cachí el día 5 marzo 2018 a las 6:00 am -----	40
Figura 23. Comparación de las 4 variables disponibles a una hora y día específicos. -----	41
Figura 24. Caso de anomalía presente en los datos, el cual es de interés para el dominio. -----	42
Figura 25. Visualización de diferentes días del año 2018 en las horas de mayor demanda energética -----	43
Figura 26. Visualización de datos energéticos de las centrales geotérmicas los días 5 de marzo y 23 de julio del año 2018. --	44
Figura 27. Predespacho Técnico Nacional correspondiente al 24.02.2018. Fuente: CENCE -----	45
Figura 28. Generación Real correspondiente al 24.02.2018. Fuente: CENCE -----	45
Figura 29. Relaciones entre las variables definidas. -----	49
Figura 30. Stream Graph.-----	51
Figura 31. Stacked Area Graph. -----	51
Figura 32. Parallel Coordinates.-----	52
Figura 33. Stream Graph: Visualización de cada una de las centrales.-----	53
Figura 34. Parallel Coordinates: Visualización de cada una de las centrales.-----	54
Figura 35. Parallel Coordinates: Datos de energía de una misma central en días determinados. -----	54
Figura 36. Stacked Area Graph: Visualización de un grupo de centrales. -----	55
Figura 37. Stream Graph: sub utilización de una central en específico. -----	56
Figura 38. Stream Graph: sub utilización de 16 centrales seleccionadas.-----	57

Figura 39. Stream Graph: sub utilización de centrales geotérmicas, biomasas y eólicas..-----	58
Figura 40. Stacked Area Graph: sub utilización de un grupo de centrales -----	58
Figura 41. Stream Graph: relación entre predespacho y energía real. -----	59
Figura 43. Stacked Area Graph: relación entre las variables. --	60
Figura 44. Stream Graph: producción de un día específico de las centrales eólicas.-----	61
Figura 45. Parallel Coordinates: producción eólica en los días 23 de julio y 10 de diciembre. -----	62
Figura 46. Stacked Area Graph: energía real generada el día 22 de julio 2018.. -----	63
Figura 47. Parallel Coordinates. -----	66
Figura 49 . Implementación final, en la cual se visualizan los datos de energía del año 2018. -----	67
Figura 50 . Técnica de brushing en los parallel coordinates.---	68
Figura 51 . Exclusión de información que no se desea visualizar. 69	
Figura 52 . Tooltip con información en detalle -----	69
Figura 53 . Visualización secundaria derivada de los Parallel Coordinates-----	70
Figura 54 . Comportamiento de las centrales eólicas durante días definidos del primer semestre del año. -----	71
Figura 55 . Tooltip que indica fecha y hora específica de las coordenadas. -----	73
Figura 56 . Comparación de las centrales Arenal y Angostura.	73
Figura 57 . Comparación de las variables de Predespacho y Generación Real.-----	74
Figura 58 . Lunes del mes de enero de las centrales eólicas. --	75
Figura 59 . Comportamiento de la producción de energía geotérmica durante diferentes días del año.-----	75
Figura 60. Producción de energía a partir de biomasa durante el año 2018.-----	76

Índice de Tablas

Tabla 1. Detalle del primer set de datos definido para el desarrollo de la visualización.-----	46
Tabla 2. Detalle del segundo set de datos definido para el desarrollo de la visualización.-----	47
Tabla 3. Detalle del tercer set de datos, muestra el total de energía real generada por día. -----	48
Tabla 4 . Matriz de evaluación según casos de uso.-----	64
Tabla 5 . Matriz de evaluación según requerimientos de visualización. -----	65

Agradecimientos

Agradezco a todas las personas que me apoyaron y facilitaron su ayuda durante el desarrollo de este proyecto.

Al profesor Franklin Hernández-Castro por compartir su conocimiento, atender todas mis dudas y hacer sesiones de gran ayuda, así como su rigurosidad y exigencia en el desarrollo del proyecto.

Gracias a mis compañeros de trabajo, en especial a Gabriela Delgado por su complicidad y apoyo incondicional durante esta etapa.

A todas las personas que conocí durante mi etapa en el TEC con las que compartí almuerzos, clases, proyectos y horas de no dormir. En especial a Ivonne Coto por su amistad, ayuda y consejos así como compartir cultura de diseño desde la primer semana de universidad.

Aunque no puedo nombrar a todos, estoy infinitamente agradecido con todos los que formaron parte de este proceso.

Dedicatoria

Dedico este proyecto a mi familia, mi papá Arturo quien fue una pieza importante para desarrollar este proyecto, y quién junto a mi mamá Isabel han dedicado tiempo desde que tengo memoria ayudándome a estudiar, hacer trabajos y desvelarse conmigo, alentándome a ser una mejor persona y profesional.

A mi hermano David por su apoyo de miles formas durante mis estudios, a quien le ha tocado ir siempre un paso adelante mío, convirtiéndose en un modelo a seguir.

Finalmente quiero dedicar este proyecto a Marcela Guzmán y Karla Araya quiénes han sido grandes mentoras compartiendo su conocimiento e impactando de gran forma en mi vida profesional motivándome a ser mejor cada día.

1. Introducción

Costa Rica se ha consolidado como líder en producción energética limpia a nivel mundial, sobrepasando a países desarrollados de la región europea y convirtiéndose en un modelo a seguir internacionalmente.

Además de los esfuerzos para convertirse en un país carbono neutral, Costa Rica se propuso para el año 2021 contar con una matriz de producción energética completamente limpia, por lo que entre el año 2014 y 2017 la producción a partir de fuentes de energía renovables aumentó en un 10%.

El sistema aprovecha 5 fuentes de energía renovables para producir energía hidroeléctrica, eólica, solar, biomásica y geotérmica; las cuáles a excepción de la geotérmica que aprovecha el calor interno de la tierra, se ven afectadas por factores externos como el clima, comportándose de forma diferente en épocas secas y lluviosas del año.

El Centro Nacional de Control de Energías, CENCE, es el ente a nivel país encargado de administrar y monitorear el Sistema Eléctrico Nacional (SEN) manteniendo la calidad del sistema y la seguridad energética en todo el territorio nacional. Para lograr esto, el CENCE envía solicitudes de generación eléctrica a todas las centrales de producción eléctrica del sistema y de estas mismas recibe datos de la generación real que fue entregada al sistema; todos estos datos de forma horaria

Actualmente el Centro Nacional de Control Energía muestra los datos relacionados a energía, pero de forma poco comprensible, por lo que la oportunidad de visualizar el comportamiento de la producción eléctrica permitirá conocer patrones entre la energía producida y la solicitud estimada así como el porcentaje de utilización de las centrales en diferentes series de tiempo.

2. Antecedentes - Indicadores

De acuerdo a un reportaje realizado por la BBC en el 2015, Costa Rica se propone que su matriz energética sea completamente limpia para el año 2021. Ante esto el panorama de producción energético a nivel nacional ha ido presentando cambios, para el año 2014 solamente el 89.68% de la energía producida proviene de fuentes renovables, 3 años después, según un informe realizado por Casa Presidencial, para el mes de diciembre del año 2017 el 99.68% de la energía generada era producto del uso de fuentes renovables.

En ese mismo año, según datos del periódico digital CRHoy, la producción energética en el país provenía de 4 principales fuentes renovables: hidroeléctrica (78.26%), eólica (10.29%), geotérmica (10.23%) y biomasa y/o solar (0.84%). Estos datos indican que el recurso hídrico se mantiene como el principal sostén de generación energética del país aumentando un 11% del 2014 al 2017.

De acuerdo a los registros del Instituto Costarricense de Electricidad (ICE), el año 2014 y 2015 fueron años muy secos y a pesar de que existía una reserva de agua en los embalses de las represas, el gobierno tuvo que recurrir al uso de otras energías como la eólica y la geotérmica, lo cual permitió el crecimiento en la infraestructura de generación de ambas.



Figura 01. Aumento de la producción energética limpia del año 2014 al 2017. Casa Presidencial 2017

El gobierno de Costa Rica realiza esfuerzos para atender la sequía que se pronostica para el año 2019 en zonas secas como lo es la región Chorotega, destinando cinco mil millones

para atender esta zona, según un informe realizado por casa presidencial en octubre del 2018. Además en declaraciones brindadas por personal del Instituto Meteorológico Nacional (IMN) al periódico La Nación, afirman que los patrones analizados del fenómeno del Niño, podrían provocar sequías en la vertiente del pacífico y más lluvias en la zona del Caribe, ocasionando una estación seca calurosa y extensa, retrasando la llegada del invierno para el 2019.

Esta situación hace pensar si ante una nueva sequía y bajos niveles de recurso hídrico la producción energética del país basada en energía hidroeléctrica será suficiente y si se está preparado al uso de los otros recursos para abastecer la demanda energética nacional.

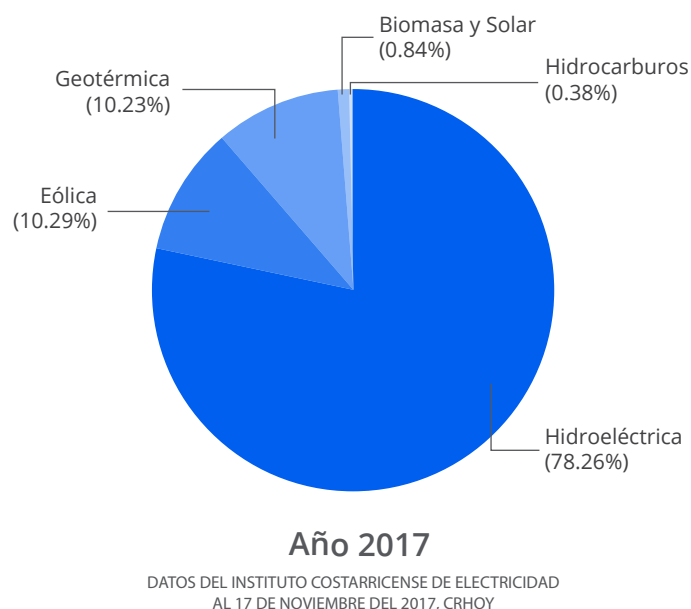


Figura 02. Distribución de la generación eléctrica nacional para el año 2017, por tipo de recurso utilizado

Para el año 2017 Costa Rica se consolidó como modelo mundial en producción energética limpia, utilizando recursos hídricos, solar, geotérmicos, biomasa y eólicas.

La energía eléctrica de Costa Rica se conecta cada vez más con las fuentes limpias: en los últimos cinco años, la participación de fuentes renovables pasaron de representar un 91,2% del total anual en el 2011, a un 98,2% en el 2016, según la recopilación de las Estadísticas de la producción de electricidad en los países del Sistema de Integración Centroamericana (SICA) difundidas por la Comisión Económica Para América Latina (CEPAL).

En cuanto a la generación pública y privada, la distancia entre ambas se recortó: la pública pasó de una participación de 85,9% en el 2011, a un 81,9% en el 2016, mientras la privada subió de un 14,1% a un 18,1% en este lapso.

La capacidad eólica instalada en el país aumentó 545% en los últimos diez años, al pasar de 70 megavatios en 2007 a 382 megavatios en 2017. Se trata de una alternativa de energía limpia que está en aumento.

El 87,43% de los 382 megavatios eólicos está en manos de productores privados.

Según Javier Orozco, director de Planificación y Desarrollo del ICE, en una entrevista para el sitio de información del Gobierno CR, la energía eólica ha venido en crecimiento en la región de Guanacaste, diversificando la matriz eléctrica de esta zona. En el 2014 solo se contaba con 4 plantas de este tipo, para el 2017 ese número ascendió a 16.

Según el reporte “Estadísticas de energía renovable 2017”, de la Agencia Internacional de Energía Renovable (IRENA), Costa Rica es el principal generador eólico de Centroamérica y el Caribe, y ocupa el tercer puesto del continente en capacidad instalada geotérmica, solamente detrás de Estados Unidos y México.

La generación más alta con este recurso se ubica en la época seca, mientras que suele reducirse en los meses de lluvias. Por ejemplo, entre marzo y mayo de 2017, su producción bajó en un 63%. Esta ausencia es suplida por los embalses del ICE (Arenal, Cachí, Pirrís, Angostura y Reventazón), que proporcionan capacidad de almacenamiento, despacho y regulación al SEN.

3. Definición del problema

3.1 Planteamiento

Los datos de producción energética recolectados por el CENCE a partir de la información brindada por la totalidad de los generadores eléctricos a nivel nacional pueden brindar una óptima compresión de dicha matriz si son visualizados de forma que se puedan determinar patrones y relaciones entre los datos.

El CENCE cuenta con una base de datos pública la cual ha sido alimentada desde el año 2008 hasta el presente, añadiendo datos diarios lo que significa un aumento considerable en el tiempo. Actualmente en su sitio web brinda algunos gráficos para analizar los datos, pero estos no son suficientes para comprender la totalidad de la información, como relaciones entre datos y patrones que se repiten a lo largo del tiempo.

La organización de los datos en el sitio web del CENCE es poco intuitiva, dificultando a los usuarios encontrarlos y visualizarlos de forma global, debido a que son presentados de forma individual.

Al ser datos públicos estos pueden ser utilizados por la comunidad científica, instituciones de comunicación y demás interesados, pero al no existir una herramienta de visualización, se dificulta su comprensión, reduciendo su valor e importancia.

3.1 Árbol del Problema

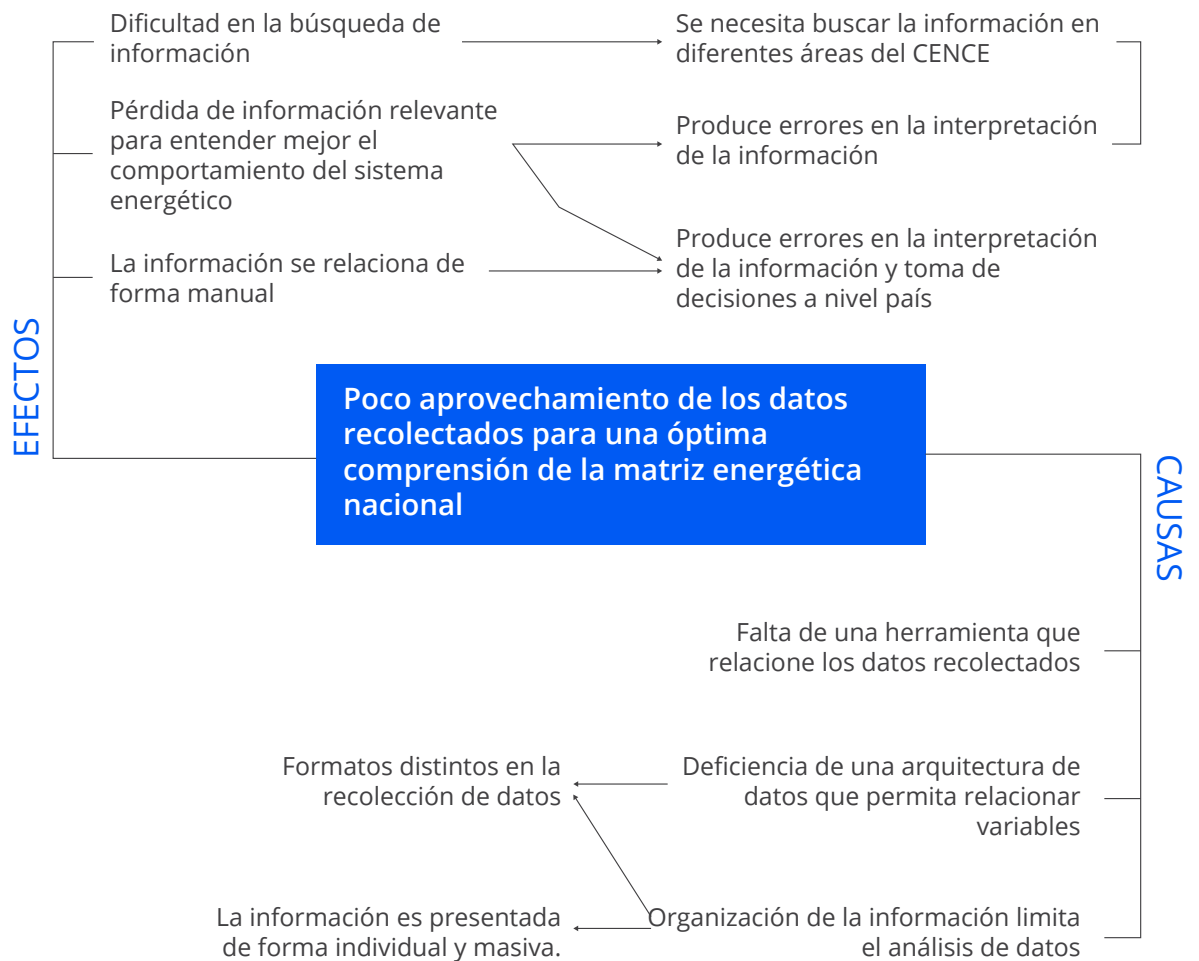


Figura 03. Árbol del Problema

4. Definición de Objetivos

Objetivo General

Diseñar una herramienta de visualización de datos que permita analizar las variables de producción eléctrica en Costa Rica a través de los datos recolectados por el CENCE, para obtener mayor valor y comprensión de estos.

Objetivos Específicos

- Analizar la naturaleza de los datos recolectados por el Centro Nacional de Control Energético, a partir de características propias.
- Determinar posibles relaciones entre los datos y las preguntas que estos pueden responder.
- Definir a partir de un análisis cualitativo las alternativas de visualización según relaciones entre datos que permitan responder las preguntas de investigación.
- Diseñar la herramienta de visualización que permita observar y analizar los datos de producción energética nacional.

5. Preguntas de Investigación

Pregunta Objetivo

¿Cuál es el grado de sub-utilización de las centrales productoras de energía?

Preguntas Secundarias

- ¿Qué época o estación del año presenta mayor grado de sub utilización de cada central eléctrica?
- ¿Cuál es la relación entre la producción solicitada y el grado de sub utilización?
- ¿Cuál es la diferencia entre la energía real generada y el grado de sub utilización?

6. Diagnóstico de la situación

6.1 Sistema Eléctrico Nacional (SEN)

Desde mediados del siglo pasado, Costa Rica explota sus fuentes renovables de manera planificada y equilibrada, mediante una matriz diversa que garantiza el suministro eléctrico del país a través de la participación pública y privada.

Este modelo único en el mundo ha permitido una cobertura eléctrica de un 99.4% de los hogares costarricenses, con una generación de más del 95% de recursos renovables. Costa Rica exhibe una matriz proveniente de recursos limpios: hídrico, geotérmico, solar, eólico y biomasa en conjunto con la generación térmica (diesel) que funciona a manera de seguro energético instalado.

Todas estas fuentes, con excepción de la geotérmica, dependen del clima. La hidroeléctrica es estacional por lo que en época seca los caudales de los ríos se reducen bajo el promedio y con la llegada de la época lluviosa sucede lo contrario. De igual forma el viento durante los meses secos aumenta su velocidad y por ende la producción energética; en los meses húmedos su velocidad disminuye.

Las restos de cosechas utilizadas para la biomásica también dependen del clima. La energía solar es variable y solo produce la mitad del día.

6.1.1 Sistema único e interconectado

Costa Rica es alimentada por un sistema inter conectado, el cual es administrado por el ICE, siendo el segundo sistema con mayor penetración en todo Latinoamérica.

Su estructura es basada en 5 fuentes renovables, en orden de volumen: agua, calor de la tierra, viento, sol y biomásica, cono un medio complementario de hidrocarburos que funciona como respaldo.

Para el 2014 la demanda energética diaria fue al rededor de 1650 megawatts, centralizado en el Gran Área Metropolitana (GAM).

6.1.2 Parques eléctricos

Los proyectos de mayor tamaño del Sistema Eléctrico Nacional (SEN) corresponde a los complejos hidroeléctricos con embalses: Arenal (Guanacaste y Alajuela), Cachí (Cartago), Angostura (Cartago), Pirris (San José) y Reventazón (Limón).

La generada por calor del suelo se agrupa en Guanacaste, por su lado la eólica se centra en las crestas montañosas del país como Guanacaste y la Zona de los Santos.

La biomásica aprovecha el bagazo de caña proveniente de ingenios azucareros, situada en el noreste del país.

6.1.3 Sistema Mixto

El grueso de la producción eléctrica se localiza en los proyectos desarrollados por el ICE desde el año 1949. Además según leyes de 1990 y 1995 el ICE debe comprar a generadores privados entre el 15% y 30% del total de la electricidad consumida por el país, pero estas empresas deben de producir con recursos renovables.

Este sistema está compuesto por 89 centrales de generación energética y para el año 2014 contaba con 444 unidades instaladas (turbinas, rotores o transformadores) de producción, la cual a aumentado con la apertura de nuevos proyectos eólicos e hídricos.

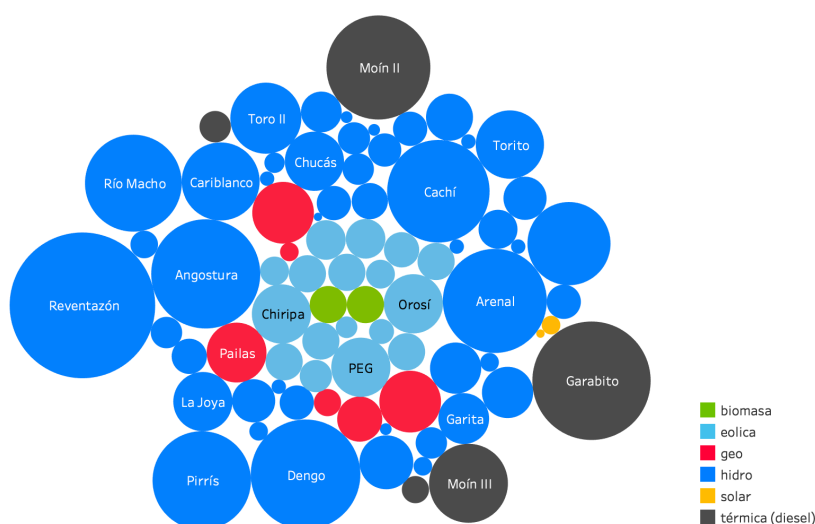


Figura 04. Sistema eléctrico nacional por central y tipo de fuente utilizada en el año 2018, según datos extraídos del CENCE.

6.1.4 Matriz Eléctrica Costarricense

La sumatoria porcentual de todos los recursos naturales de los que se extrae la energía para transformarla en electricidad y llevarla a hogares, comercios e industrias corresponde a la matriz eléctrica costarricense.

En Costa Rica esta matriz es administrada y monitoreada por el ICE, mediante el Centro Nacional de Control de Energía (CENCE) y el Sistema Eléctrico Nacional (SEN), facilitando la tarea de mantener la calidad del sistema, además de vigilar y garantizar la seguridad eléctrica en todo el territorio nacional.

6.2 Generación de Energía

La energía producida por las centrales en Guanacaste es consumida en diferentes lugares como Puntarenas y Limón por ejemplo, esto significa que la interconexión del sistema garantiza el suministro a cualquier parte del país.

Los involucrados en el proceso de generación de energía se clasifican en 4 tipos: ICE, Empresas (CNFL, cooperativas, entre otros), Privados y BOT (built-operate-transfer); este último se licita siguiendo la ley, para que un desarrollador privado se encargue del proyecto, lo explote y lo entregue al ICE al cabo de un periodo establecido de tiempo.

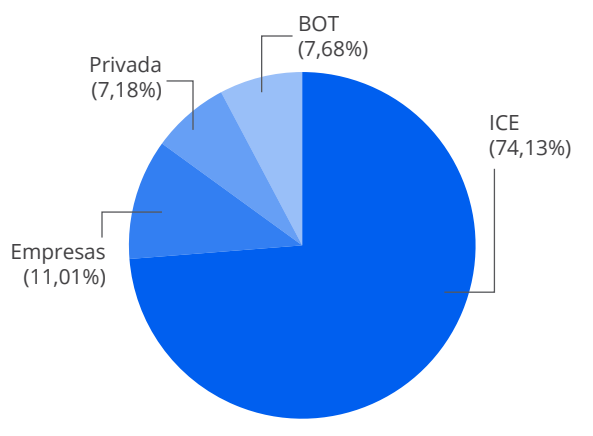


Figura 05. Porcentaje de participación según tipo de entidad productora. Datos al 31 de diciembre del 2014.

Para el año 2014 se mapeó el tipo de recurso renovable utilizado por cada productor para la generación de energía, como se observa a continuación.

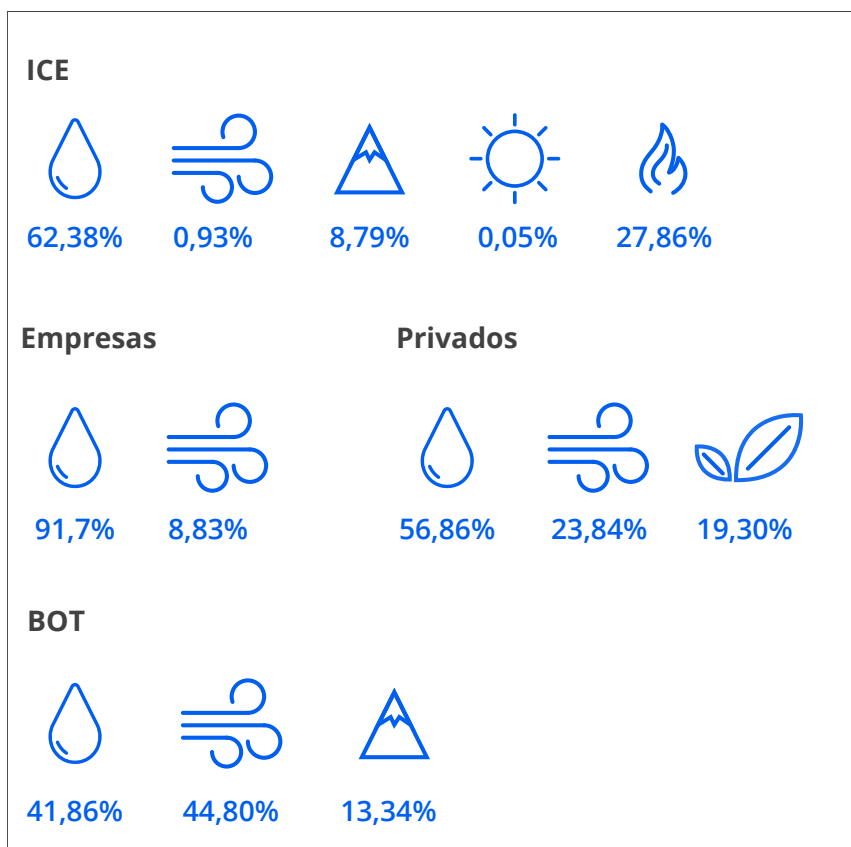


Figura 06. Capacidad instalada según tipo de recurso utilizado por los diferentes productores.

6.2.1 Planificación de la producción energética

El CENCE como ente administrador de la matriz energética se encarga de regular la producción de todos los involucrados. Para esto genera una solicitud estimada de energía en megawatts por hora (mW/h) a producir por cada central generadora a través de una distribución horaria (24 horas). Dicha solicitud se conoce como **Predespacho Técnico Nacional** y es enviada a cada generador un día antes de que se efectúe la generación real.

Una vez que el predespacho es enviado, la central energética se prepara para producir dicha solicitud, la cual puede variar por diferentes factores como el clima, disponibilidad de unidades o de fuente renovable. Esta generación se conoce

como Generación Real y responde directamente a la demanda energética del país.

Además dentro de la generación real se encuentran las variables de Intercambio, que es toda aquella energía que entra o sale de la matriz energética nacional a través de las fronteras norte y sur. Este intercambio puede ser negativo, correspondiendo a toda la energía que salió, o puede ser positivo, significando que energía externa fue aportada al sistema nacional.

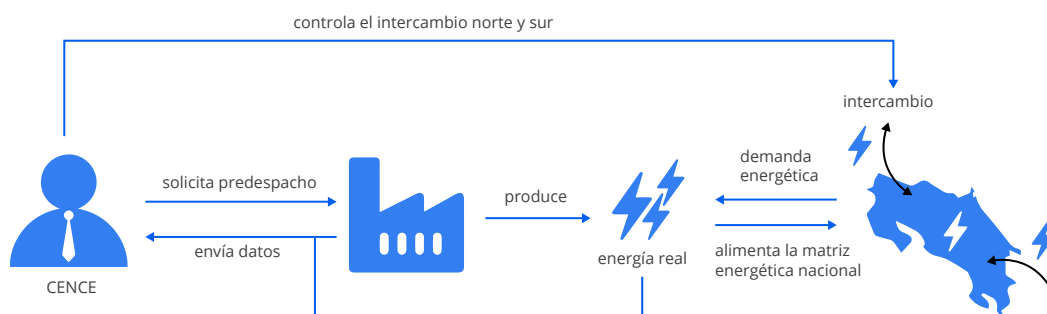


Figura 07. Flujo de planificación y producción energética entre el CENCE y productores del mercado.

Es común que la solicitud de producción realizada a través del predespacho técnico nacional sea diferente a la energía real producida por la incidencia de distintos factores. En algunos casos corresponde a una gran diferencia al visualizar los datos por medio de tablas, es por esto que los datos de generación real recolectados por el CENCE deberían permitir realizar una solicitud estimada mas precisa entre predespacho y la energía que se generará realmente, comprendiendo mejor el comportamiento de la matriz energética.

Este análisis de los datos podría permitir tomar mejores decisiones, conocer si la captación de los datos es óptima, comportamientos y patrones que no son obvios, así como conocer la inactividad o la actividad de las unidades por factores internos o externos al sistema.

7. Marco Teórico

7.1 La Generación de Energía

7.1.1 Matriz Eléctrica

Es la sumatoria porcentual de todos los recursos naturales de los que se extrae la energía para transformarla en electricidad y llevarla a hogares, comercios e industrias (referencia folleto matriz ICE)

7.1.2 Capacidad Instalada

Potencia nominal o de placa de una unidad generadora, o bien se puede referir a una central, un sistema local o un sistema inter conectado.

7.1.3 Predespacho Técnico Nacional

Modelo matemático general que permite caracterizar los sistemas energéticos y realizar una predicción de producción de toda la red energética.

7.1.4 Energías Limpias

La energía limpia es un sistema de producción de energía con exclusión de cualquier contaminación o la gestión mediante la que nos deshacemos de todos los

residuos peligrosos para nuestro planeta. Las energías limpias son, entonces, aquellas que no generan residuos.

7.1.5 Central generadora

Lugar y conjunto de instalaciones utilizadas para la producción de energía eléctrica. Dependiendo del medio utilizado para producir dicha energía, recibe el nombre correspondiente

7.1.6 Central hidroeléctrica

Central generadora que produce energía eléctrica utilizando turbinas que aprovechan la energía potencial y cinética del agua.

7.1.7 Central eólica

Central generadora que produce energía eléctrica utilizando turbinas que aprovechan la energía cinética del viento.

7.1.8 Central geotérmica

Central generadora que produce energía eléctrica utilizando turbinas que aprovechan la energía calorífica del vapor de agua, producido a lo interno de la tierra.

7.1.9 Central solar

Instalación industrial en la que, a partir del calentamiento de un fluido mediante radiación solar y su uso en un ciclo termodinámico convencional, se produce la potencia necesaria para mover un alternador para generación de energía eléctrica

7.1.10 Factor de Planta

Conocido también como factor de utilización de una central, es la relación entre la energía eléctrica producida por un generador o conjunto de generadores, durante un intervalo de tiempo determinado y la energía que habría sido producida si este generador o conjunto de generadores hubiese funcionado durante ese intervalo de tiempo, a su potencia máxima posible en servicio. Se expresa generalmente en porcentaje.

7.1.11 CENCE

Centro Nacional de Control de Energías, entidad encargada de controlar el sistema y red energética de Costa Rica.

7.1.12 Clima

Conjunto de condiciones atmosféricas propias de un lugar, constituido por la cantidad y frecuencia de lluvias, la humedad, la temperatura, los vientos, etc., y cuya acción compleja influye en la existencia de los seres sometidos a ella.

7.2 Visualización de Datos (DataViz)

Corresponde a una visualización de datos e información, la cual permite obtener conocimiento a profundidad de los datos analizados de forma comprensible para el usuario. Se utiliza como una herramienta para ayudar al análisis y no como sustitución de este.

7.2.1 Big Data

El Big Data se refiere a gran cantidad de datos específicos que no pueden ser procesados o analizados utilizando herramientas o procesos tradicionales. Estos datos provienen de una acumulación masiva de datos provenientes de empresas, industrias, investigaciones, propios de la generación del conocimiento humano.

7.2.2 Paradigmas de Visualización

Diferentes tipos de visualización de los datos conocidos como paradigmas, corresponden a taxonomías de los datos que pueden ir desde gráficos de pastel, diagrama de circos, grandes y complejas matrices de los mismos.

Algunos autores clasifican los datos en 6 dimensiones, mientras que otros los clasifican por el nivel de tareas.interacción, representación y análisis. Según (Hernández-Castro y Monge-Fallas, 2016) la mayoría de las visualizaciones se dividen en 4 categorías principales, “node-links”, “adjacency”, “enclosure” y “networks”. Dentro de estas categorías podemos encontrar diferentes estilos de representación como sunburst, treemaps, chords, entre otros.

Dependiendo de lo que se quiera exponer con los datos algunas de estas representaciones son más óptimas que otras, lo cual dependerá de las características de estos, como locación, atributos, relaciones y tiempo.

Con la finalidad de dirigir la presente investigación a una visualización óptima según la naturaleza de los datos obtenidos diariamente por el CENSE, se decidió seguir la taxonomía de The Data Visualisation Catalogue propuesta por Severino Ribbecca de London College of Communication específicamente para visualizar datos en períodos de tiempo y relacionales (Ribbecca, s.f.).

Algunos de los paradigmas propuestos en The Data Visualization Catalogue se presentan a continuación.:

Stacked Area Graph

Trabajan del mismo modo que los gráficos de área, pero en este caso utiliza múltiples series de datos.

El gráfico completo representa el total de los datos, transmite números enteros por lo que no funciona para valores negativos. Son útiles para comparar múltiples variables que cambian en un intervalo de tiempo.

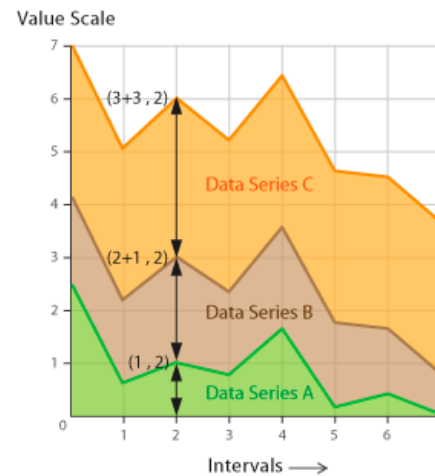


Figura 08. Stacked Area Graph.
Fuente: The Data Visualisation Catalogue.

Bubble Chart

Gráfico de multi variables que utiliza un sistema de coordenadas cartesianas, colocando puntos a lo largo de una cuadrícula donde el eje X y Y son variables separadas. Cada burbuja representa una tercer variable dependiendo del área del círculo.

Comúnmente utilizados para comparar y mostrar relaciones utilizando la posición y proporción. La visualización completa permite analizar patrones y correlaciones.

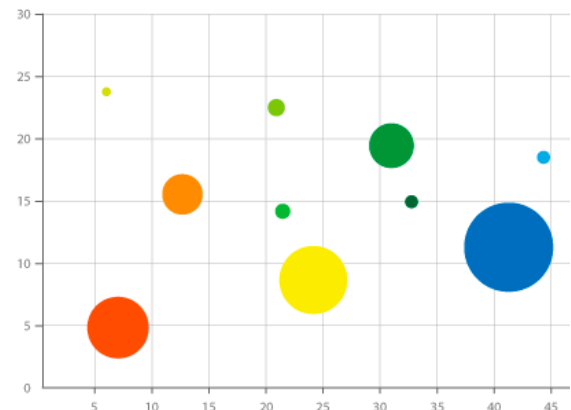


Figura 09. Bubble Chart. Fuente: The Data Visualisation Catalogue

Heatmap

Visualizan datos a través de una variación de color. Útiles para cruzar datos multivariable, mostrando patrones, correlaciones entre datos y similitudes.

Generalmente todas las columnas son una categoría y las líneas otra, subdivididas en sub categorías.

Debido a su dependencia del color para comunicar valores son útiles para obtener una visión general de los datos pero dificultan la posibilidad de extraer información específica de estos.

Se utilizan para mostrar cambios a través del tiempo, por ejemplo compara cambios de temperatura en un periodo anual en diferentes ciudades.

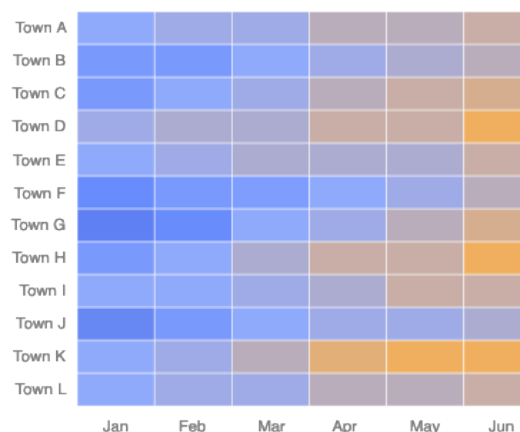


Figura 10. Heatmap. Fuente: The Data Visualisation Catalogue.

Histograma

Visualiza la distribución de datos a través de un intervalo continuo de tiempo. Cada barra representa la frecuencia del dato.

Permite tener un estimado de dónde están concentrados los datos así como probabilidades de distribución.

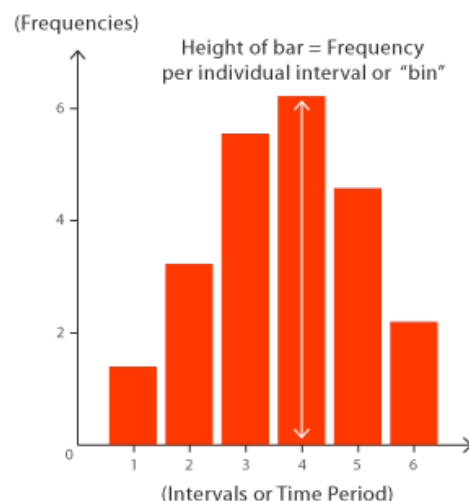
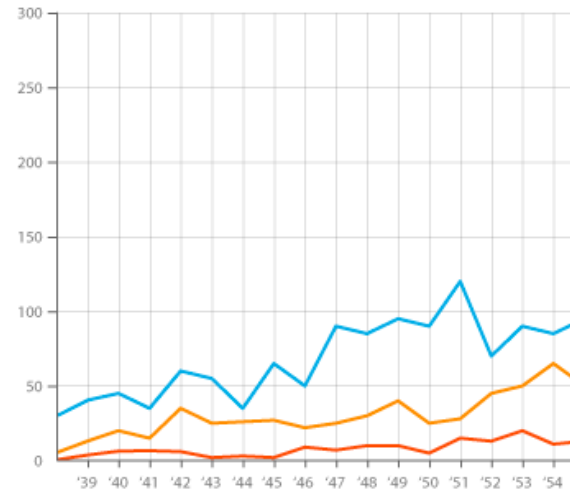


Figura 11. Histograma. Fuente: The Data Visualisation Catalogue, Severino Ribecca.

Gráfico de líneas

Utilizado para mostrar datos cuantitativos a través de un intervalo continuo de tiempo, mostrando tendencias y cambios a través del tiempo.

Cuando las líneas son agrupadas con otras permite hacer comparaciones, aunque gran cantidad de líneas aumenta la dificultad de lectura de la visualización.



aumento desproporcionado de valor.

Figura 12. Gráfico de líneas. Fuente: *The Data Visualisation Catalogue*.

Nightingale Rose Chart / CoxComb Chart

Utilizado por la médico y estadística Florence Nightingale para comunicar las causas de muerte de soldados durante la guerra de Crimea.

Dibujado sobre coordenadas polares, donde cada categoría esta dividida en segmentos iguales, que tan largo se extiende el segmento del centro, depende del valor que representa.

Una desventaja es que el aumento de área de los segmentos externos les brinda mas énfasis y un

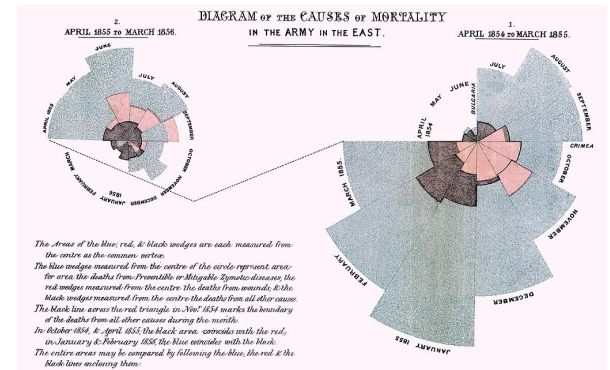


Figura 13. Causas de mortalidad del ejército del este en la guerra de Crimea. Autora: Florence Nightingale.

Spiral Plot / Time Series Spiral

Traza datos de tiempo a lo largo de una espiral arquimediana. Inicia en el centro de la espiral y avanza hacia el exterior.

Son muy versátiles debido a que permiten usar barras, líneas o puntos a lo largo del camino de la espiral.

Ideal para mostrar grandes conjuntos de datos y tendencias en un periodo prolongado de tiempo.

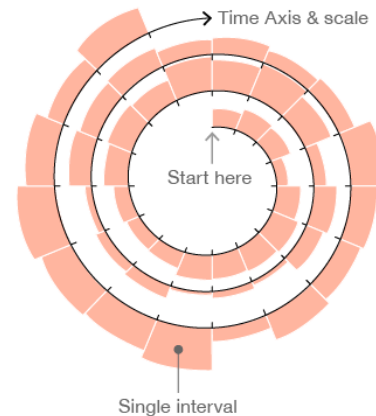


Figura 14. Spiral Plot. Fuente: The Data Visualisation Catalogue.

Gráfico de Flujo / Stream Graph

Gráfico de flujo con valores desplazados alrededor de una línea central. Muestra cambios en los datos a lo largo del tiempo de diferentes categorías mediante el uso de formas orgánicas fluidas., esto hace que sea estéticamente agradable.

Ideal para mostrar conjuntos de datos de gran volumen, con el fin de descubrir tendencias y patrones a lo largo del tiempo en una amplia gama de categorías.

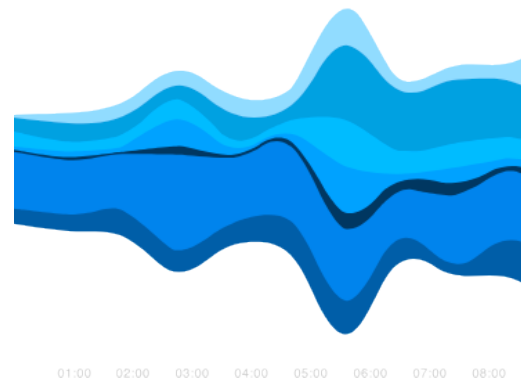


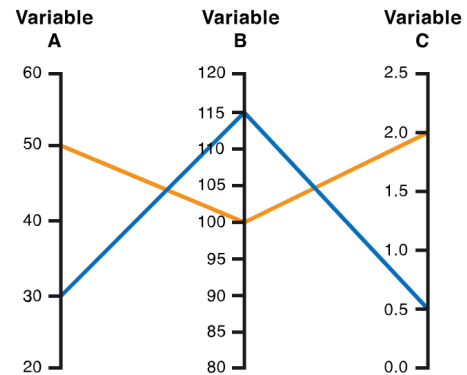
Figura 15. Gráfico de Flujo. Fuente: The Data Visualisation Catalogue.

Brinda una visión general pero dificulta ver datos específicos en casos de gran granularidad.

Parallel Coordinates

Este tipo de visualización es utilizada para comparar datos numéricos multi-variables. Es ideal para comparar muchas variables juntas y observar las relaciones entre ellas.

Cada variable tiene su propio eje y es dispuesta de forma paralela una de otra. Cada eje puede tener una escala diferente o compartir una sola. Es decir cada línea es una colección de puntos.



*Figura 16. Parallel Coordinates.
Fuente: The Data Visualisation Catalogue, Severino Ribecca.*

Chord Diagram

Este tipo de diagrama visualiza las relaciones entre diferentes entidades. Las conexiones entre estas entidades muestran que comparten algo en común. Son ideales para comparar similitudes entre los sets de datos.

Las entidades están ubicadas en una circunferencia, utilizando arcos o curvas de Bézier para interconectar estas entidades.

Se complica su lectura cuando hay muchas conexiones.

Parallel Sets

Muestran flujo y proporciones a través de líneas de flujo, las cuales corresponden a una dimensión del set de datos y son divididas por líneas que indican las categorías.

Cada línea de flujo tiene un color que muestra y compara la distribución ante otras categorías.

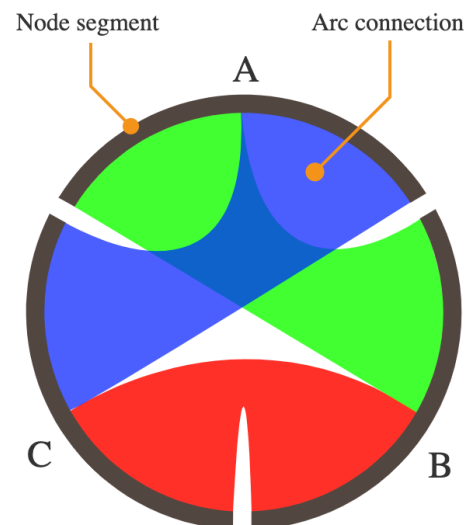


Figura 19. Chord Diagram. Fuente: The Data Visualisation Catalogue.

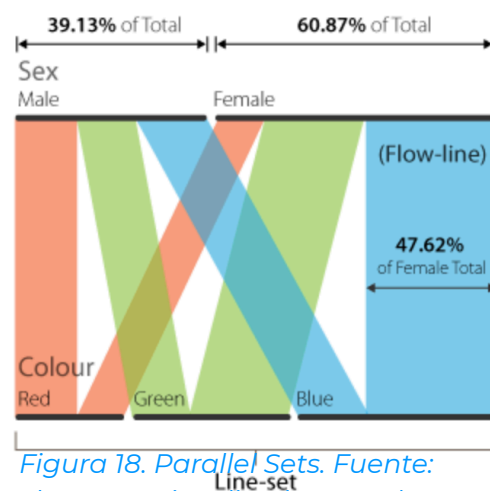


Figura 18. Parallel Sets. Fuente: The Data Visualisation Catalogue.

8.Marco Metodológico

8.1 Investigación

El trabajo desarrollado corresponde a una investigación cuantitativa, a partir de una necesidad encontrada de visualizar de una mejor manera diferentes aspectos y datos de la red eléctrica nacional, utilizando una visualización de datos. Para esto se delimita la necesidad, el problema, definiendo también la guía metodológica de investigación y desarrollo.

Además, se toma en cuenta la experiencia de diferentes involucrados, específicamente generadores eléctricos privados, así como las estadísticas, datos e informes del Centro Nacional de Control de Energías (CENSE), permitiendo comprender de forma óptima los datos y sus variables.

A partir de este desarrollo metodológico se podrá obtener un resultado acorde a los objetivos planteados y brindar una herramienta para poder comprender y analizar a profundidad la gran

granularidad de datos que poseen los involucrados.

Recolección de Datos

La principal base de datos para la presente investigación son de uso público y se encuentran disponibles en el portal en línea del Centro Nacional de Control de Energías - CENCE, en el siguiente enlace: <https://apps.grupoice.com/CenceWeb/CenceMain.jsf>

8.2 Metodología

8.2.1 Selección de Tema

En esta etapa se selecciona un tema y necesidad que se pueda resolver a partir de una visualización de datos.

8.2.2 Investigación Preliminar

Se investiga qué se ha estudiado y qué esfuerzos se han realizado acerca del tema y su problemática.

8.2.3 Recolección de Datos

Consiste en la recolección de datos que pueden ser utilizados para dar una solución a la problemática a través de una visualización de datos.

8.2.4 Análisis de relaciones entre datos

Se toman los datos recolectados para analizar relaciones no obvias entre estos, de forma que permitan obtener un conocimiento a profundidad del tema.

8.2.5 Preguntas de análisis

En esta etapa se proponen posibles preguntas de las cuales se puede obtener respuesta a través de la visualización de datos. Se define una pregunta objetivo y secundarias que apoyadas en las relaciones de datos estudiados.

8.2.6 Data Cleaning

Durante esta etapa se limpian los datos y se organizan para poder ser utilizados en el desarrollo de la visualización.

8.2.7 Preselección de Paradigma de Visualización

A partir de las relaciones, la pregunta objetivo y la limpieza de datos, se realiza un análisis acerca de los tipos

de visualización óptimos para poder dar respuesta a la necesidad y problema.

8.2.8 Análisis y evaluación de los paradigmas

Para la selección de los paradigmas propuestos, se realiza una matriz de selección basada en casos de uso (según preguntas que se desean responder) y criterios de diseño y usabilidad (carga cognitiva, facilidad de uso, eficiencia y eficacia). A partir de estos se realiza una puntuación y el paradigma con la puntuación más alta será el seleccionado para implementación.

8.2.9 Implementación

8.2.10 Validación

Se valida con potenciales usuarios la visualización a partir de tareas y casos de uso, determinando así si la calidad de los análisis que pueden realizarse, recomendaciones y trabajos a futuro.

10. Análisis de Referenciales

10.1 Referenciales

Con la finalidad de analizar elementos positivo de visualizaciones de referencia, se revisaron diferentes tipos de visualizaciones de datos temporales y utilizarlos

10.1.1 The daily routines of famous creative people

Combina una cantidad considerable de datos en una sola página, utilizando el color para distinguir tendencias y al mismo tiempo permite al usuario observar de forma global la totalidad de información e interactuando con ella para obtener datos o contenido mas específico.

Utiliza el eje x como medida de tiempo y en el eje y muestra grupos de datos, en este caso por nombre de la persona analizada.

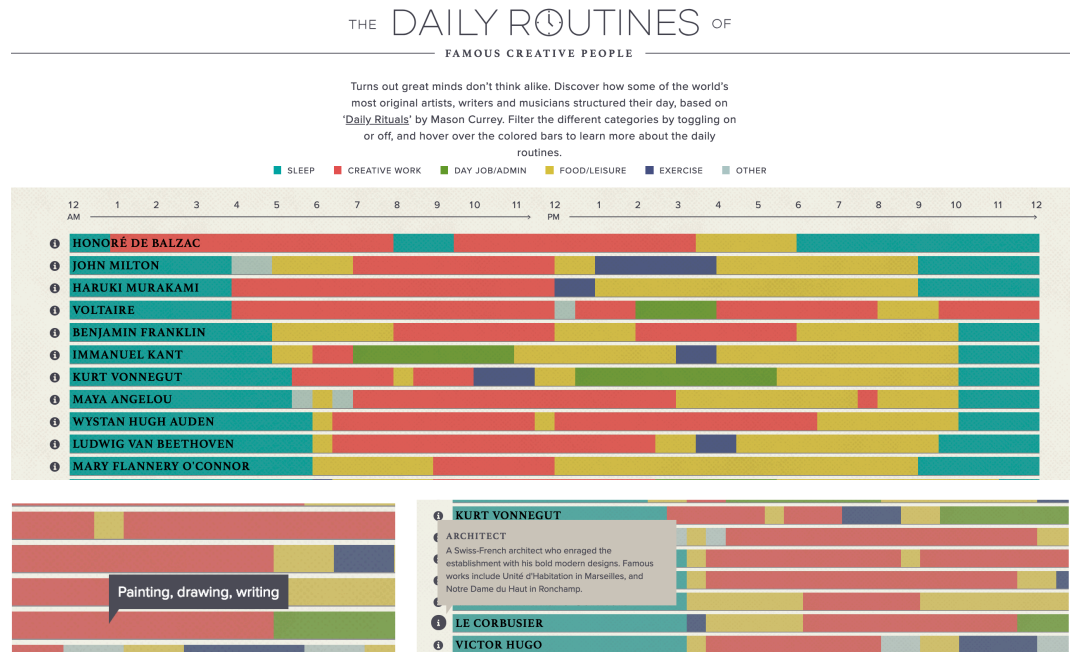


Figura 19. Bar Graph. The daily routines of famous creative people.
Fuente:Podio

10.1.2 What's really warming the world

La finalidad de esta visualización es desaprobar diferentes teorías de que el calentamiento global se está dando por causas naturales. Muestran como la temperatura ha ido elevándose desde 1880 hasta el presente, el eje X corresponde a una línea del tiempo y el eje Y a datos de temperatura.

Se utiliza un gráfico de líneas para representar las diferentes causas (el sol, volcanes, cambios en la órbita, deforestación, ozono, aerosoles y gases invernaderos), permitiendo comparar todos los factores, por grupos, separados y promedios.

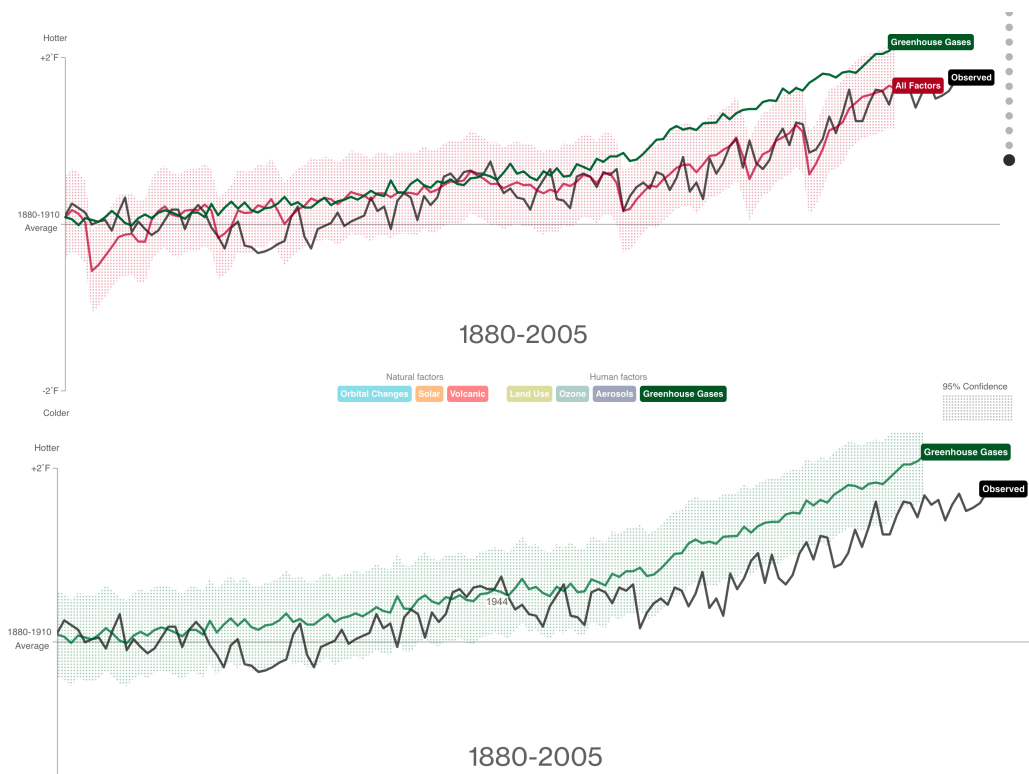


Figura 20. Line Graph. What's really warming the world?. Fuente: *Bloombergs Business Work*, 2015.

10.1.3 The complete history of the NFL

Visualiza datos de resultados de cada uno de los juegos de toda la historia de la Liga de Fútbol Americano, correspondientes a mas de 30 000 datos. El usuario puede comparar cada equipo y observar como ha sido su desempeño a través de décadas de juego.

Utiliza gráficos de línea que permiten visualizar tendencias, picos de buen o mal desempeño a través de una línea de tiempo la cual puede ser controlada por un scrollbar teniendo la opción de ver un período determinado o la totalidad de la historia.

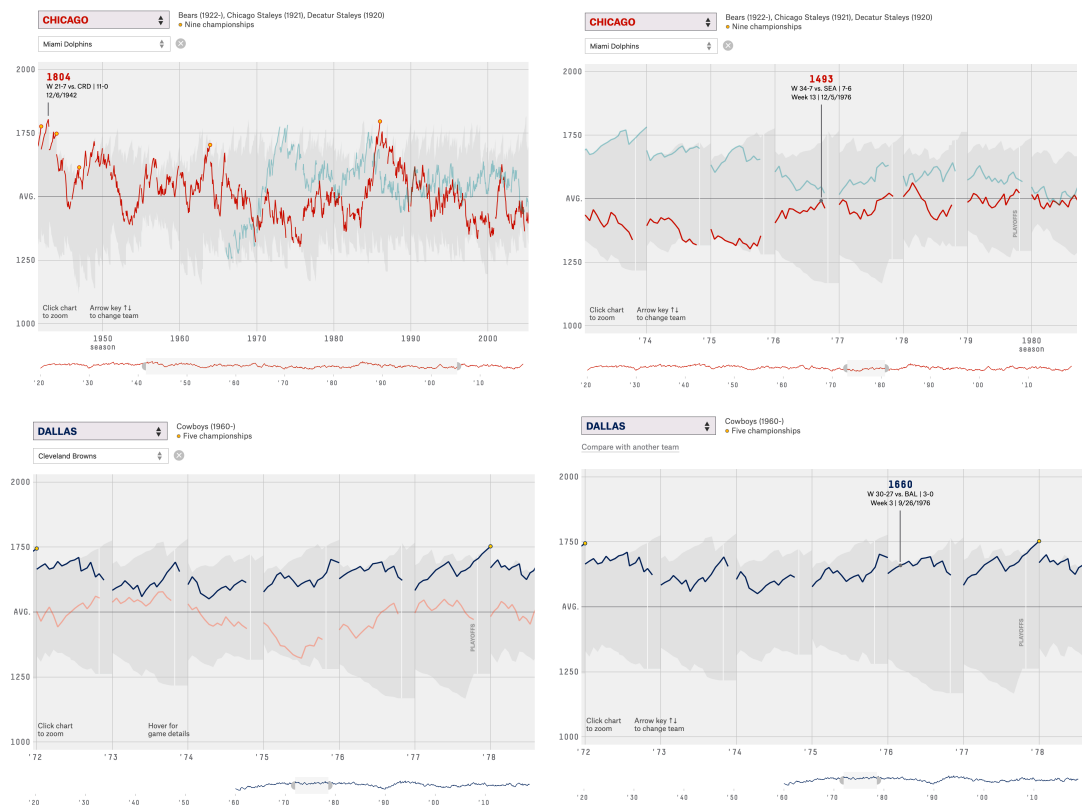


Figura 21. Line Graph. The Complete History of the NFL. Fuente: Project Five Thirty Eight.

10.1.4 The year in news

Echelon Insights analiza los temas que las personas hablaron en twitter en el año 2014, a través de una visualización de Stream Graph que permite ver tendencias y los temas que fueron mas “twiteados”, además de poder visualizar las noticias que fueron mas compartidas en periodos de tiempo semanales.

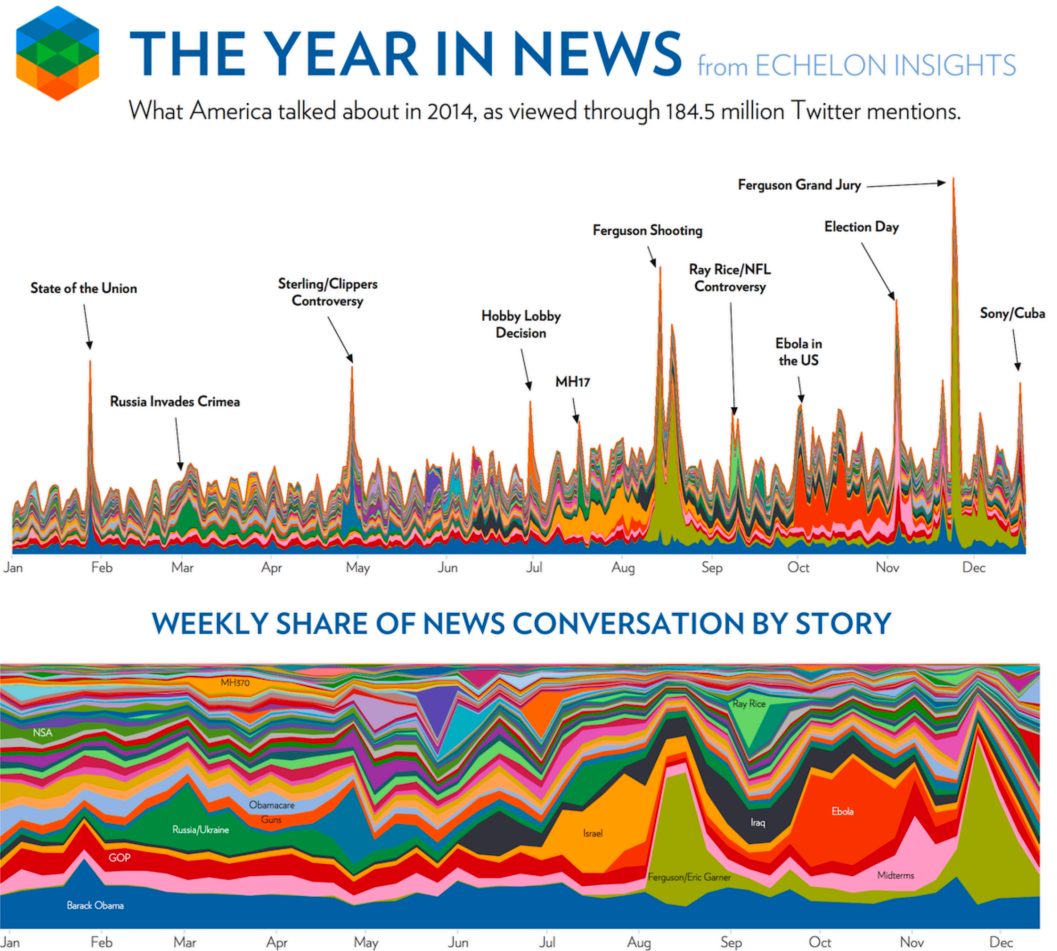


Figura 22. Stream Graph. The Year in News. Fuente: Echelon Insights, 2015.

10.1.5 Terrorist attacks

Analiza la causa de ataques terroristas desde el año 1998 hasta el 2015 con la finalidad de observar momentos históricos así como tendencias entre los datos, para esto se utiliza un Stream Graph que permite visualizar el río de información a través de la línea del tiempo, de forma agradable y estética.

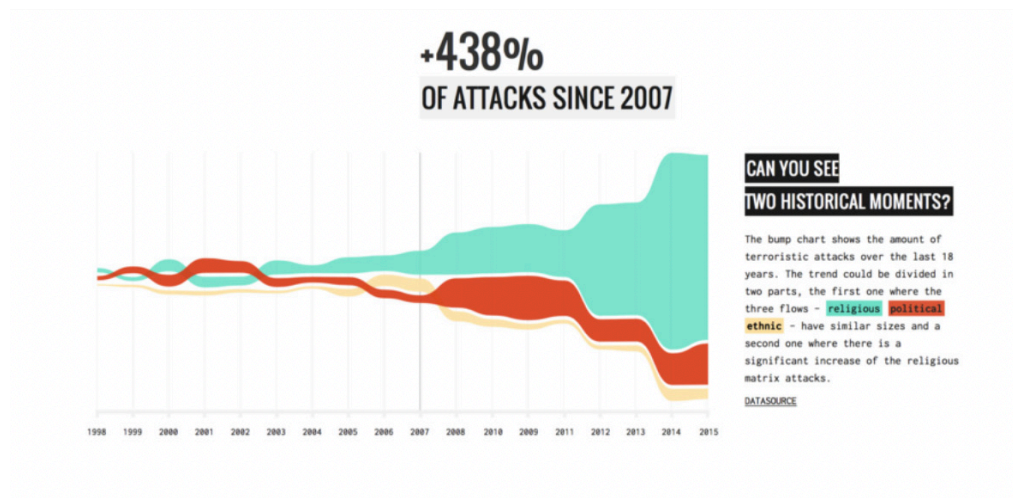


Figura 21. Bump Chart. Terrorist Attacks. Fuente: Density Design

10.2 Conclusiones y mínimos comunes.

A partir del análisis realizado se definen aquellos elementos en común que presentan los referenciales con la finalidad de determinar un punto de partida para el proyecto a desarrollar.

Se determina que la visualización deberá:

- Ser simple
- Diferenciación por colores
- Utilizar una línea de tiempo debido a la naturaleza propia de los datos.
- Brindar un “big picture” de los datos al usuario, así como un acercamiento a partes específicas.
- Brindar información adicional al usuario si desea conocer más del dato.
- Comparar datos o grupos de datos.

11. Análisis de los datos

Los datos recolectados por el CENSE corresponden a la **generación real**, los cuales se encuentran distribuidos en formato horario empezando a las 00:00 horas y finalizando a las 23:00, por cada una de las centrales productoras que comprenden el parque eléctrico nacional. Estos datos horarios son recolectados de forma diaria.

De igual forma el CENSE diariamente envía los datos de **predespacho**, los cuales mantienen la misma configuración que los datos recolectados.

La totalidad de los datos se encuentra dividida por central generadora y los datos de energía se encuentran en kilowatt por hora (kW-h) y megawatt por hora (mW-h).

Estos datos han sido recolectados desde el año 2008 hasta la actualidad, actualizándose de forma diaria.

Para comprender mejor cómo se distribuyen los datos se toma como ejemplo la Planta Hidroeléctrica de Cachí cuya capacidad instalada es de 152 000 kW-h. El día 5 marzo del 2018 para las 6:00 am los datos de predespacho enviados por el CENSE correspondieron a 32 500 kW-h y la energía real producida fue de 49 426 kW-h.



Figura 22. Generación real VS predespacho de la planta hidroeléctrica Cachí el día 5 marzo 2018 a las 6:00 am

Es importante destacar que actualmente las comparaciones de estos datos de energía solo se pueden hacer de forma manual, ya que se ubican en diferentes tablas y documentos del dominio.

11.1 Desarrollo de Hipótesis

Con la finalidad de comprender mejor los datos, se plantearon diferentes hipótesis que permitieran visualizar información específica. Estas hipótesis se muestran a continuación.

11.1.1 Visualizar todas las variables a una hora específica

Se tomó una hora y un día específico de forma que se pudiera visualizar el comportamiento de cada una de las variables de Capacidad Instalada, Predespacho y Energía real producida.

Además se realizó una diferencia entre la capacidad instalada y energía real visualizando **la capacidad instalada que no está siendo utilizada**, de color rojo en la figura 22.

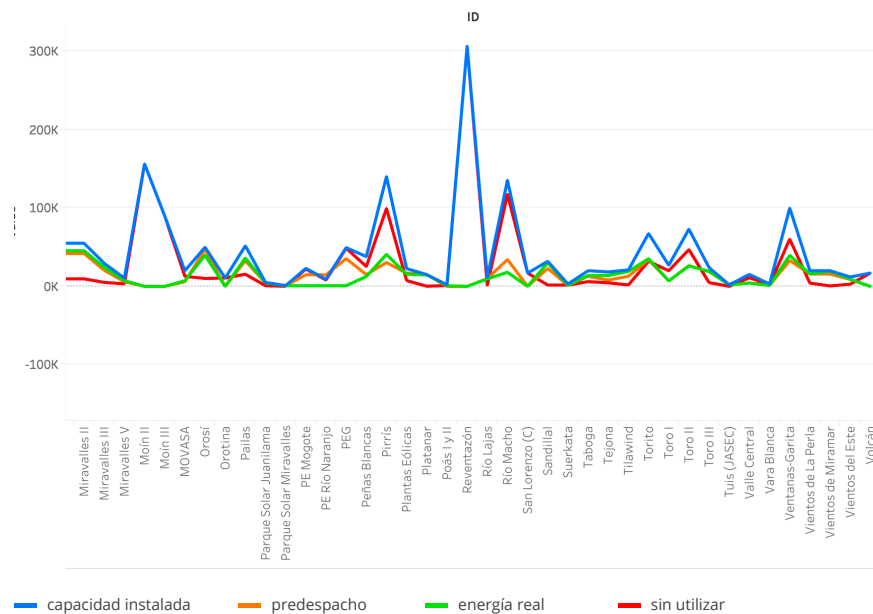


Figura 23. Comparación de las 4 variables disponibles a una hora y día específicos.

Esta visualización permitió ver anomalías en los datos, por ejemplo, como se muestra en la figura 24, las centrales energéticas Doña Juana y Don Pedro registraron un valor de energía sin utilizar mayor a su capacidad instalada y a la energía real generada a esa hora específica, lo cual de acuerdo al dominio no puede ser verdadero, por lo que se toma como una anomalía al momento de registrar o recolectar los datos.

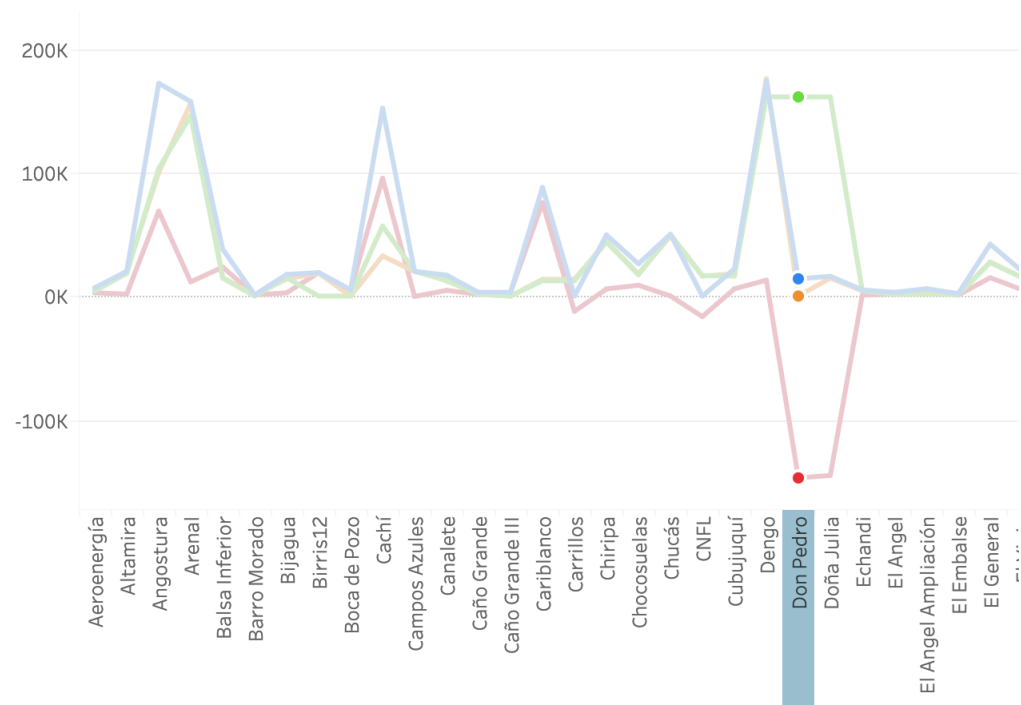


Figura 24. Caso de anomalía presente en los datos, el cual es de interés para el dominio.

11.1.2 Comparar el mismo día de la semana en diferentes meses y horas

Para esto se toma una muestra de cuatro lunes de diferentes semanas con la finalidad de analizar comportamientos y patrones, de forma lineal.

La muestra elegida corresponde a los días 5 marzo, 23 julio, 24 setiembre y 10 diciembre, pertenecientes al año 2018 y en dos horas de alta demanda de electricidad, específicamente las 6:00 am y 12:00 pm.

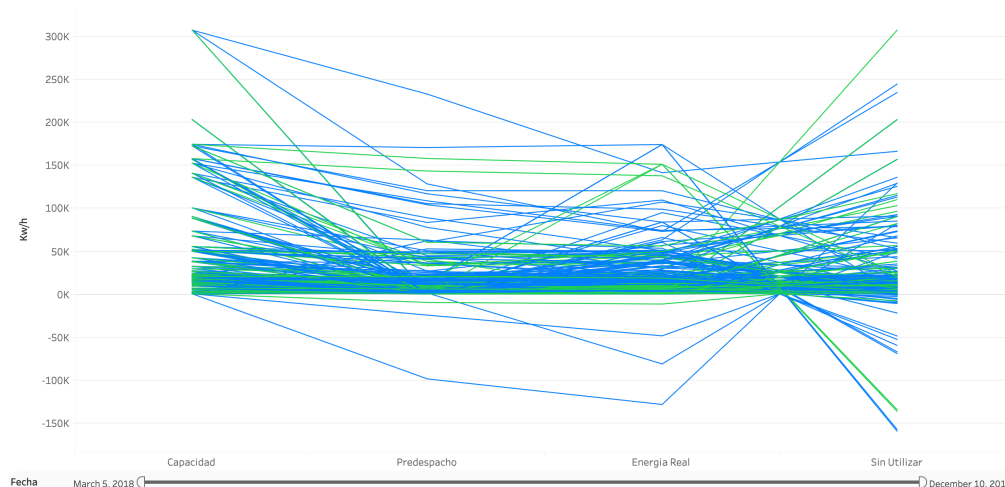


Figura 25. Visualización de diferentes días del año 2018 en las horas de mayor demanda energética

Los colores de las líneas corresponden a la hora, siendo las 6:00 am verde y las 12:00 pm azul, de forma que se puede entender que la demanda energética es mayor al medio día, además de visualizar cuales son las centrales mas utilizadas y aquellas que están siendo sub-utilizadas.

A través del uso de un slider para controlar la cantidad de días que se quieren visualizar, así como el tipo de energía y la hora se pueden visualizar datos específicos y no de forma global, permitiendo comprender de mejor forma los datos.

Por ejemplo, en la figura 26, se puede visualizar la producción energética de las centrales geotérmicas en los días 5 de marzo y 23 de julio, a las 6 am y 12 pm, obteniendo así información a detalle.



Figura 26. Visualización de datos energéticos de las centrales geotérmicas los días 5 de marzo y 23 de julio del año 2018.

Una vez que se han realizado las hipótesis y se comprenden los datos y su naturaleza de mejor forma, se procede a realizar el proceso de data mining y data cleaning.

11.2 Data Mining

Como se mencionó al inicio de esta sección, se utilizan los set de datos disponibles en la base de datos pública del CENCE, los cuáles pertenecen a la información diaria recolectada, divididos en dos grandes grupos.

El primero de estos corresponde a los datos de **Predespacho Técnico Nacional**, organizados de forma horaria en sus columnas y en sus filas muestra el nombre de la central eléctrica a la que corresponden los datos energéticos (ver figura 27).

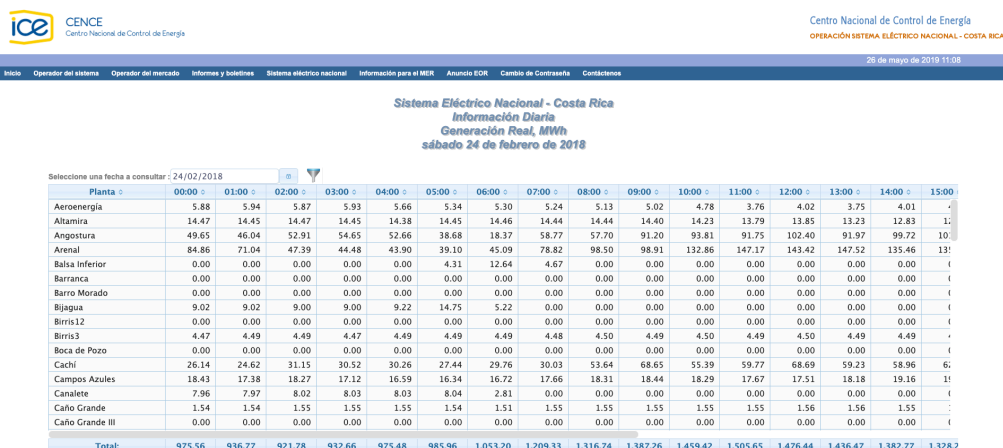


Figura 27. Predespacho Técnico Nacional correspondiente al 24.02.2018. Fuente: CENCE

La **Generación Real** de cada una de las centrales conforma el segundo set de datos, estos están organizados con la misma estructura del set anterior (ver figura 28).

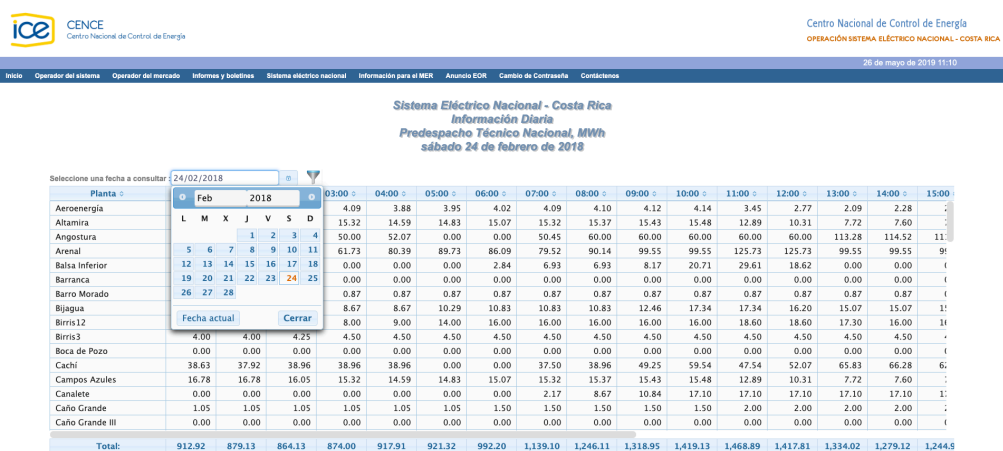


Figura 28. Generación Real correspondiente al 24.02.2018. Fuente: CENCE

Ambos grupos de datos son organizados en hojas de cálculo diarias, de forma que para realizar una consulta se debe buscar en el filtro de calendario para seleccionar el día y año deseado, por lo tanto cada día desde el año 2009 posee su propia hoja de cálculo.

Esta distribución de la información dificulta la comparación de datos energéticos en un periodo de tiempo definido. Además el sitio web presenta gráficos poco intuitivos que no permiten una óptima comparación de las variables.

El tercer set de datos está conformado por información proveniente de diferentes fuentes, de forma que se puede conocer el nombre de cada una de las centrales, su capacidad instalada y el tipo de fuente que utilizan para generar energía (eólica, hidroeléctrica, geotérmica, entre otros).

11.3 Data Cleaning

Una vez definidos los datos a utilizar para el desarrollo de la visualización, se procede a realizar la limpieza de datos (data cleaning). Para esto se unifican los datos recolectados en dos sets de datos.

Definición de los Set de Datos

El primer set de datos definido corresponde a la unión de datos de predespacho técnico nacional y generación real, los cuales están directamente relacionados, a continuación se muestra la estructura del nuevo data set.

Fecha	Hora	ID	Predespacho	Energia Real	No Generada
01/01/2018	12:00	Aeroenergía	4930	3722.6	3,027.40
01/01/2018	12:00	Altamira	12280	20117.5	-117.50
01/01/2018	12:00	Angostura	131250	0.0	172,202.00
01/01/2018	12:00	Arenal	104940	29728.0	127,671.00
01/01/2018	12:00	Balsa Inferior	33000	32595.8	5,404.20
01/01/2018	12:00	Barranca	0	0.0	53,280.00
01/01/2018	12:00	Barro Morado	660	0.0	1,000.00
01/01/2018	12:00	Bijagua	14590	17873.7	-373.70

Tabla 1. Detalle del primer set de datos definido para el desarrollo de la visualización.

En la tabla 1 se puede apreciar la nueva estructura de los datos, la primer y segunda columna corresponden a datos de tiempo, específicamente fecha y hora. La tercer columna llamada ID, indica el nombre de la central eléctrica. A partir de la cuarta columna se encuentran datos energéticos en kilowatts por hora (kW/h), correspondientes a predespacho y energía real. La sexta y última columna (destacada en color azul) corresponde a la diferencia entre la capacidad instalada de cada un de las centrales y la energía real producida, obteniendo datos de energía no generada (capacidad instalada sin utilizar).

El segundo set de datos definido para realizar la visualización se aprecia en la Tabla 2. La estructura de este set está conformada por tres columnas, la primera indica el nombre de cada una de las centrales eléctricas, lo cual permite correlacionarse con el primer set de datos. El tipo de fuente utilizada para la producción eléctrica por cada una de las centrales se indica en la segunda columna. Por último la tercer columna corresponde a la capacidad instalada que posee cada una de las centrales para producir energía, estos datos se encuentran en kW/h.

ID	Fuente	Capacidad Instalada
Aeroenergía	eolica	6,750.00
Altamira	eolica	20,000.00
Angostura	hidro	172,202.00
Arenal	hidro	157,399.00
Balsa Inferior	hidro	38,000.00
Barranca	térmica (diesel)	53,280.00
Barro Morado	hidro	1,000.00
Bijagua	hidro	17,500.00
Birris12	hidro	19,000.00
Birris3	hidro	13,600.00
Boca de Pozo	geo	5,000.00

Tabla 2. Detalle del segundo set de datos definido para el desarrollo de la visualización.

El tercer y último set de datos establecido corresponde a la producción total diaria de cada una de las centrales energéticas, la cual por su estructura puede relacionarse con los sets de datos establecido anteriormente. A continuación en la Tabla 3 se observa la estructura y los datos que este set posee.

La primer columna corresponde al nombre de la central eléctrica (ID), las siguientes columnas corresponden al total de la generación real de cada una de las centrales por cada día del año.

ID	Enero 1	Enero 2	Enero 3	Enero 4	Enero 5
Aeroenergía	71880.1	124492.1	122188.2	149831.9	136264.9
Altamira	299687.6	480038.6	457965	427372.8	436734.4
Angostura	34996	740812	889556.5	771092	685244
Arenal	792608	1335808	1231264	1425376	1107584
Balsa Inferior	279644	247934.5	316809.7	339564.4	300970.3
Barranca	0	0	0	0	0
Barro Morado	0	0	0	0	0
Bijagua	309073.5	367331.7	352501.7	291318.1	292554.9
Birris12	234574.4	221752.6	288627.3	260859.6	298986.3
Birris3	104145.8	102976.9	102607.2	107326.6	98715
Boca de Pozo	0	0	0	0	0
Cachí	319074.75	670997.25	416348.25	407115.5	61496.5

Tabla 3. Detalle del tercer set de datos, muestra el total de energía real generada por día.

Dimensión de las variables

Al definir y aclarar la estructura de los datos, se establecen las relaciones entre ellos así como las principales dimensiones, correspondientes a:

- **ID:** nombre de cada una de las 88 centrales eléctricas analizadas.

- **Fuente:** el tipo de recurso renovable o no renovable utilizado por cada central para la producción de electricidad.
- **Capacidad Instalada:** capacidad de producción eléctrica de cada central si esta opera al 100% de su límite establecido.
- **Fecha y Hora:** corresponden a datos de tiempo necesarios para someter a comparación los datos energéticos.
- **Predespacho:** solicitud de producción eléctrica realizada por el CENCE a cada una de las centrales, en kilowatts por hora.
- **Generación Real:** energía real producida por cada central, datos en kilowatts por hora.
- **Energía no generada:** información en kilowatts por hora obtenida de la diferencia entre capacidad instalada y generación real, es decir la capacidad instalada sin utilizar en cada una de las centrales.

Al definir estas dimensiones es clara la naturaleza de los datos, dividida en dos tipos, **series de tiempo y datos categóricos**.

Las relaciones entre las variables definidas se pueden apreciar en la figura 29.

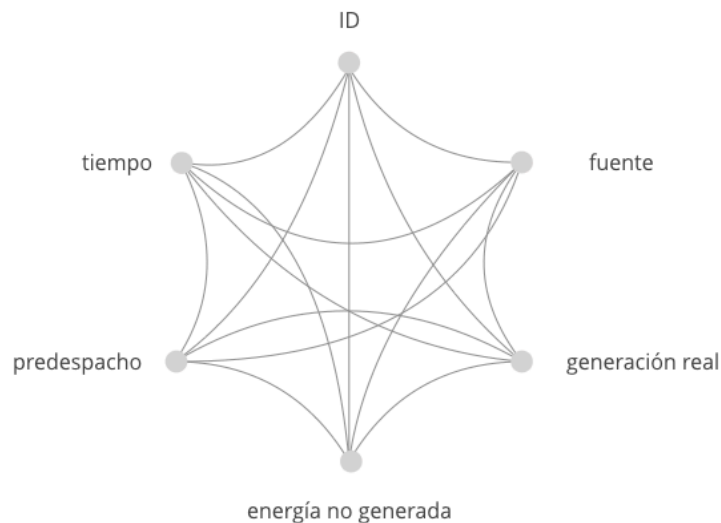


Figura 29. Relaciones entre las variables definidas.

12. Paradigmas de visualización

Finalizado el análisis y limpieza de los datos, se procede a realizar la selección de los paradigmas de visualización que permiten observar las relaciones entre los datos según su naturaleza y que ayudan a responder las preguntas objetivos planteadas.

Para esto se toma en cuenta que se deben visualizar las relaciones entre 4 variables (capacidad instalada, predespacho, energía real y capacidad sin utilizar), ya sea de forma global o de forma específica.

Además se debe de poder visualizar el tipo de fuente que se utiliza para generar la energía y el nombre de la central energética.

Por otro lado estos datos deben poder ser visualizados de forma diaria y horaria.

A continuación se muestran los paradigmas seleccionados y analizados.

12.1 Paradigmas Seleccionados

12.1.1 Stream Graph

Se definió la visualización de río o Stream Graph, el cual permite ser manipulado para visualizar la información global, detalles y comparaciones. Cada línea de flujo representa una central energética y su forma corresponde al comportamiento en el tiempo.

En el eje vertical se encuentra un rango de energía en kW/h, el cual permite determinar la cantidad de energía producida o solicitada de cada central. El periodo de tiempo, en este caso horario lo encontramos en el eje horizontal.

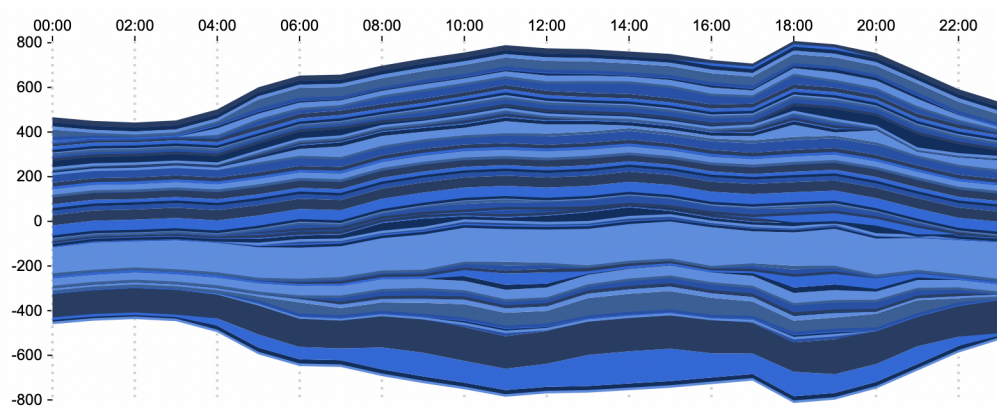


Figura 30. Stream Graph.

12.2 Stacked Area Graph

Se propone el uso de esta visualización debido a la versatilidad de visualizar múltiples series de datos. Cada área representa a una central eléctrica en un período de tiempo definido y la totalidad del gráfico muestra los datos globales. Esta visualización permite comparar las variables de forma simultánea.

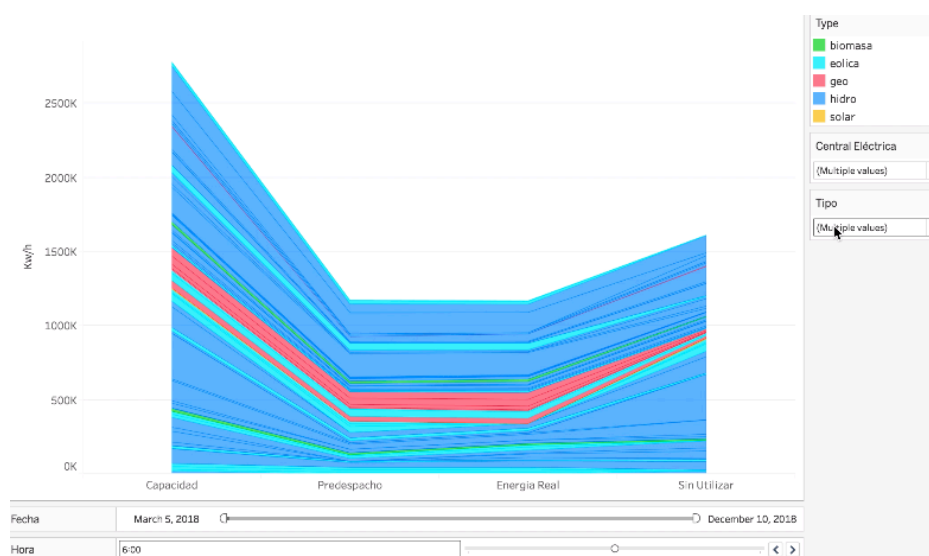


Figura 31. Stacked Area Graph.

En el eje horizontal se muestran 4 categorías de clasificación y el eje vertical muestra un rango de producción eléctrica en kiloWatt por hora (kW/h). A través de una línea de tiempo interactiva, las horas y días pueden ser determinadas por el usuario. El color es dado por el tipo de fuente renovable utilizada.

12.3 Parallel Coordinates

Se definió una estructura lineal para visualizar las cuatro variables correspondientes a datos categóricos. Cada una de estas variables se encuentran en el eje horizontal, formando columnas a través de las cuales se distribuye la información.

En el eje vertical se encuentra un rango de energía en kW/h, el cual permite determinar la cantidad de energía de cada una de las columnas. El periodo de tiempo, horario y diario, se puede determinar por medio de una línea de tiempo interactiva. Al igual que el Stacked Area Graph, el color es dado por el tipo de fuente renovable utilizada.

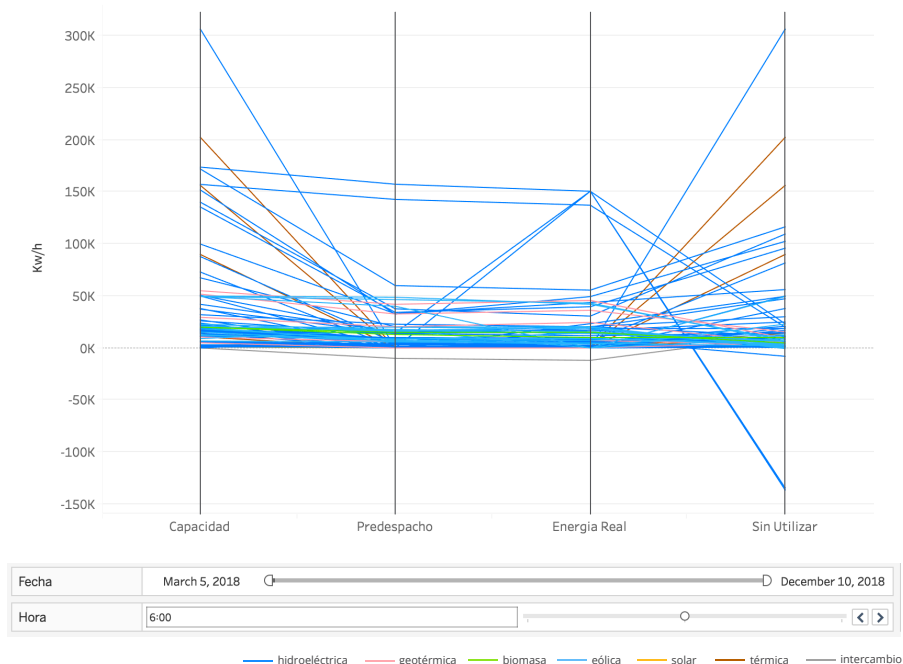


Figura 32. Parallel Coordinates.

12.2 Casos de Estudio

12.2.1 Visualizar cada central eléctrica

Stream Graph

El usuario tiene la posibilidad de realizar mouse over y mouse click sobre cada uno de los flujos, mostrando un tooltip con información de ID, hora y datos de producción energética, relacionado a una variable.

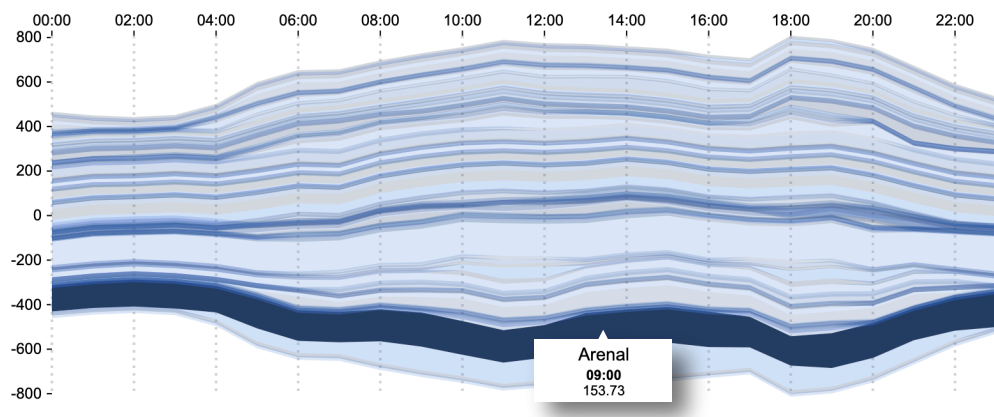


Figura 33. Stream Graph: Visualización de cada una de las centrales.

Parallel Coordinates

El usuario el usuario puede seleccionar la central y el tipo de fuente para visualizar la información de producción de la central, obteniendo información de las 4 variables analizadas a través de un tooltip, el cual muestra el ID, hora, tipo de fuente y el dato exacto de producción en la coordenada seleccionada.

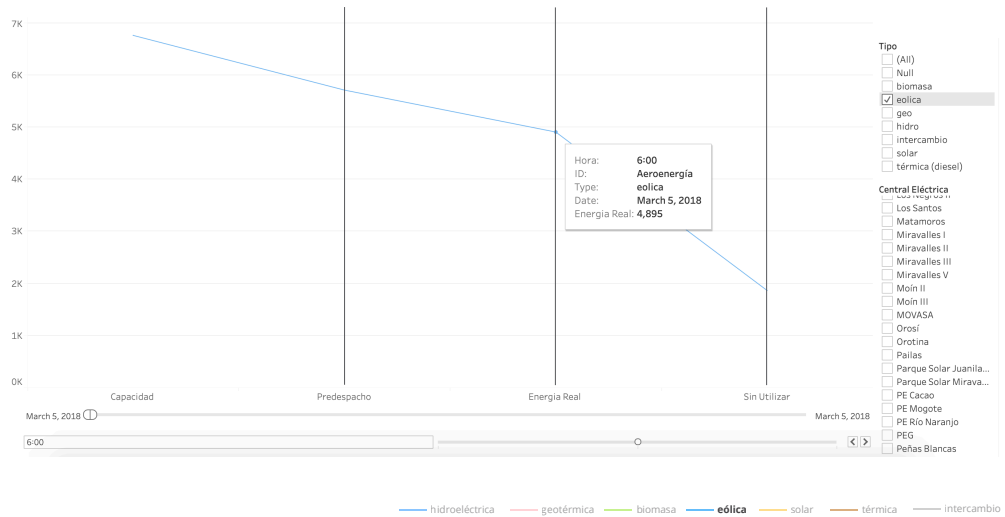


Figura 34. Parallel Coordinates: Visualización de cada una de las centrales.

Los parallel coordinates permiten visualizar el comportamiento de una central eléctrica en específico a través de un periodo de tiempo. En este caso cada línea corresponde a un día de producción de la central visualizada.

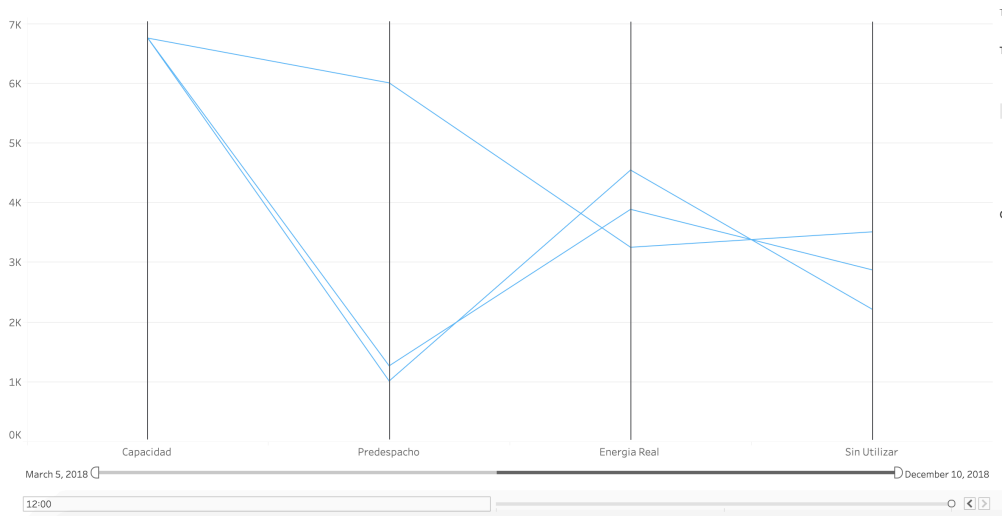


Figura 35. Parallel Coordinates: Datos de energía de una misma central en días determinados.

Stacked Area Graph

Permite observar cada área haciendo focus sobre la sección seleccionada, como se muestra en la figura 36 por medio de un tooltip muestra la información relacionada a la variables, así como el ID y datos de tiempo.

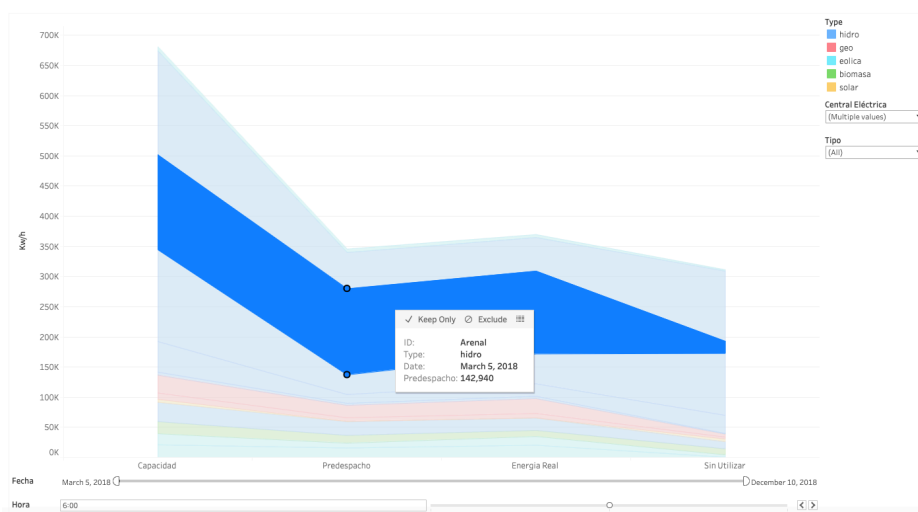


Figura 36. Stacked Area Graph: Visualización de un grupo de centrales.

Si el usuario desea puede excluir el resto de centrales y solo observa una en específico, o bien comparar un grupo de estas como se muestra en la figura 27.

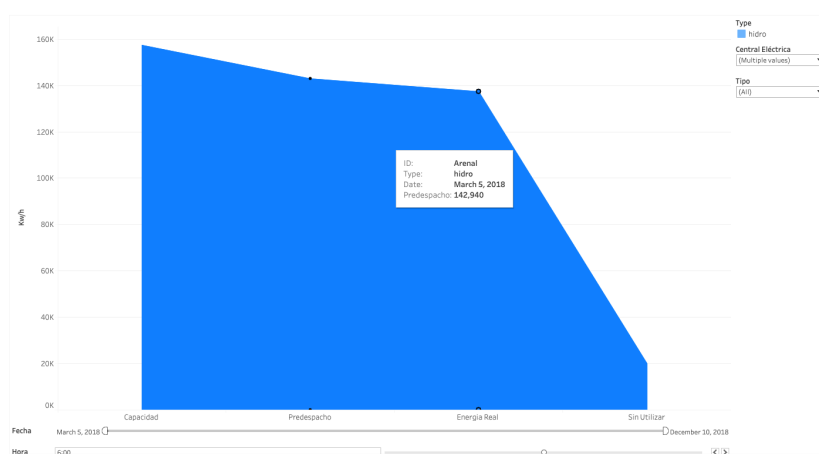


Figura 36. Stacked Area Graph: Visualización de una central.

12.2.2 Subutilización de las centrales

Stream graph

Permite visualizar cada planta por separado, obteniendo una comparación entre las 4 variables (predespacho, energía real, capacidad y capacidad sin utilizar), observando la diferencia energética entre cada una.

En el caso de la figura 37, se observa el comportamiento de la Central Hidroeléctrica de Arenal, donde tamaño en el eje vertical corresponde al dato de energía y flujo o “río” se mueve a través de 24 horas

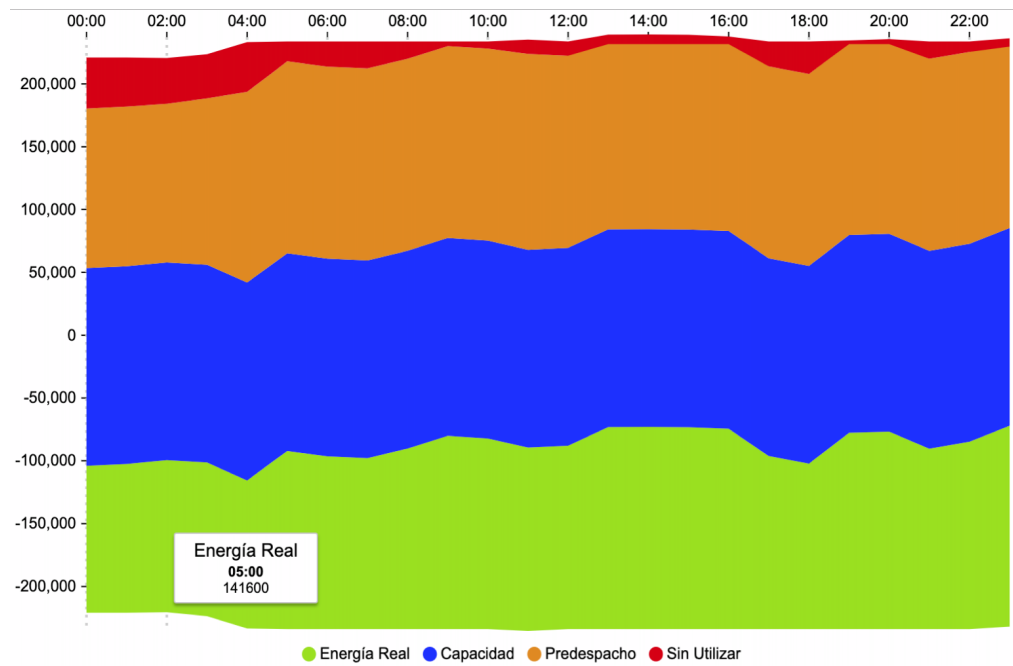


Figura 37. Stream Graph: sub utilización de una central en específico.

Además de visualizar una sola central energética, se puede realizar una comparación entre la sub utilización de toda la matriz eléctrica nacional o un conjunto de estas.

En el caso de la figura 38, se tomaron 16 centrales para visualizar el comportamiento de la sub utilización en un día completo.

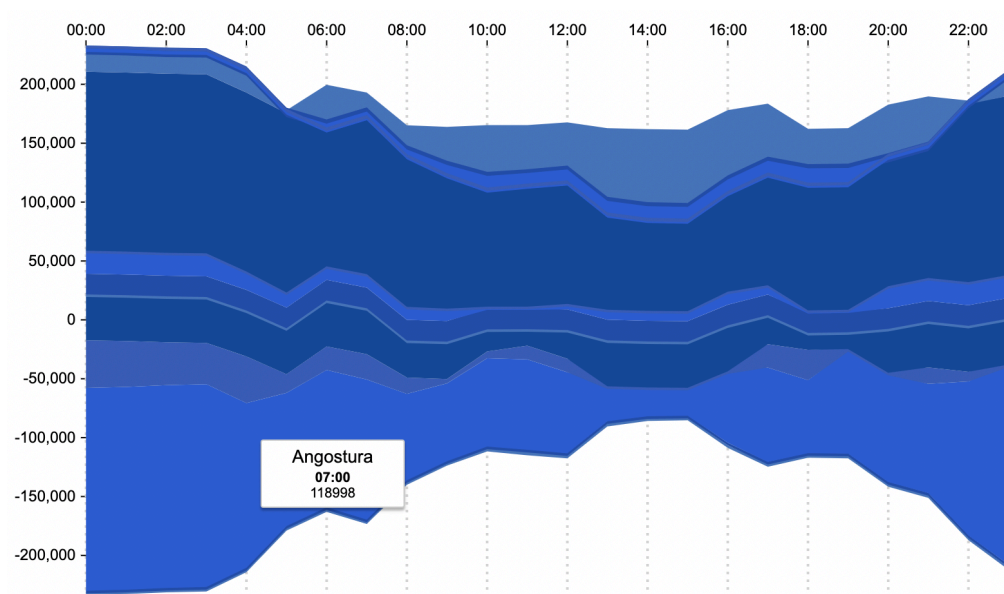
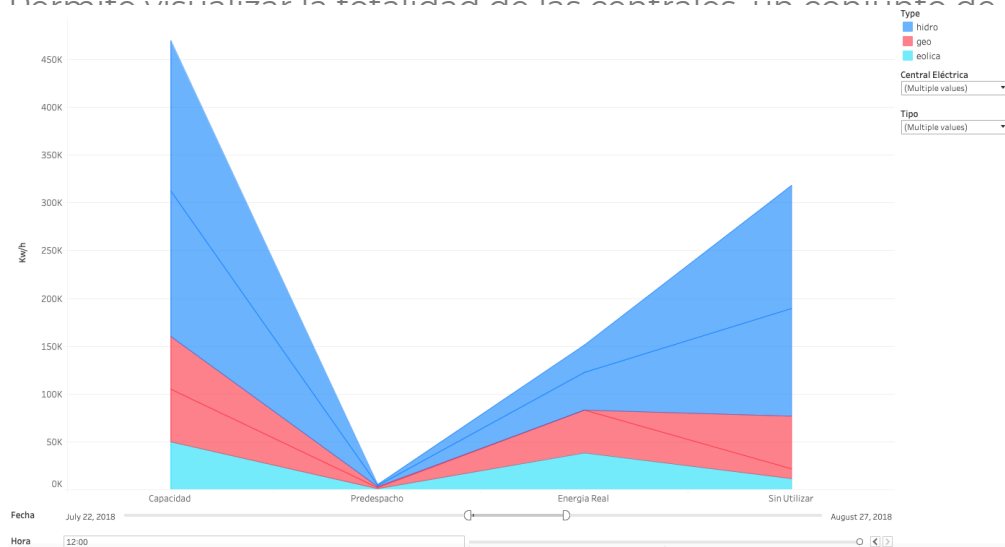


Figura 38. Stream Graph: sub utilización de 16 centrales seleccionadas.

Parallel Coordinates

Permite visualizar la totalidad de las centrales en un conjunto de



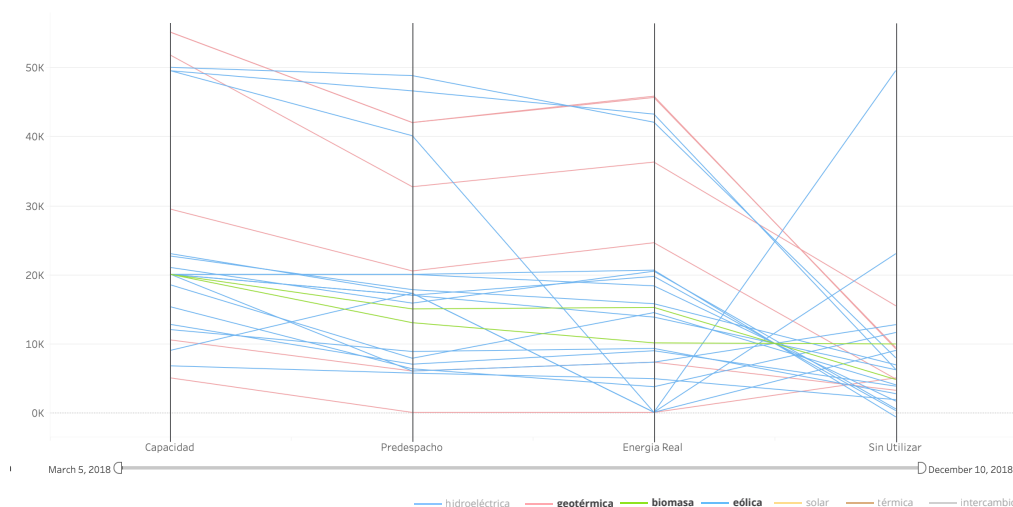


Figura 39. Stream Graph: sub utilización de centrales geotérmicas, biomasa y eólicas..

Stacked Area Graph

En el stacked area graph es observable la energía no generada o capacidad sin utilizar de cada una de las centrales que el usuario haya decidido mostrar, además permite realizar comparaciones al mismo tiempo con respecto a las demás variables.

Figura 40. Stacked Area Graph: sub utilización de un grupo de centrales

12.2.3 Relación entre predespacho y energía real

Stream graph

Permite seleccionar la central que se desea analizar, observando la relación entre la energía real y el predespacho, obteniendo la curva de cada variable durante un periodo de tiempo establecido en el eje horizontal.

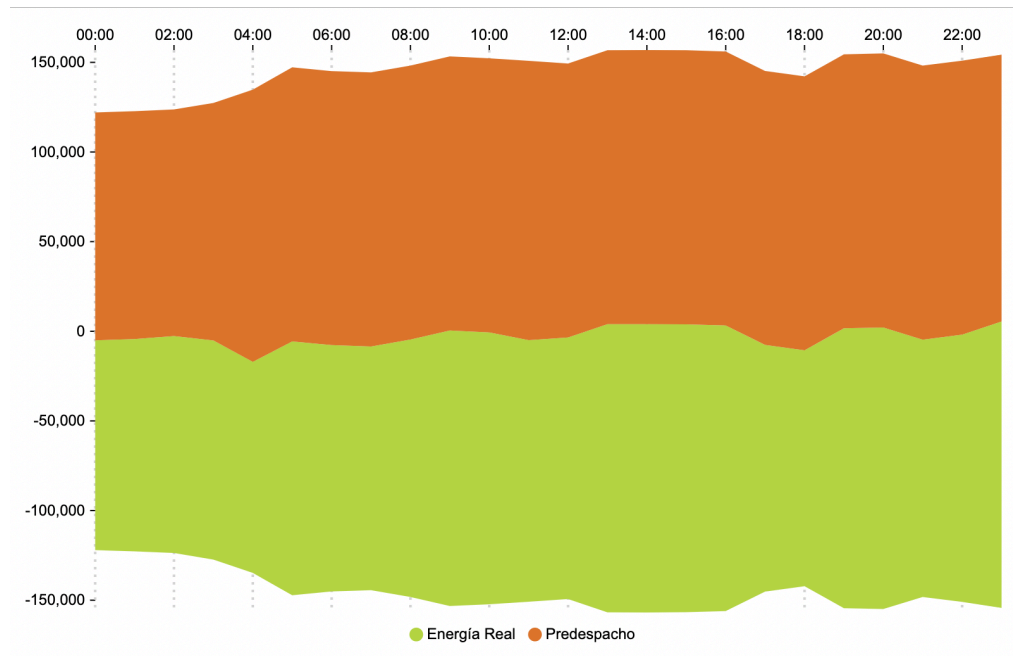


Figura 41. Stream Graph: relación entre predespacho y energía real.

Parallel Coordinates

El usuario puede seleccionar las centrales y el tipo de fuente que desea analizar y correlacionar, obteniendo patrones y diferencias entre las variables de predespacho y energía real, en periodos de tiempo definidos. Además, indirectamente se observa cuales son las centrales que poseen una mayor capacidad productiva.

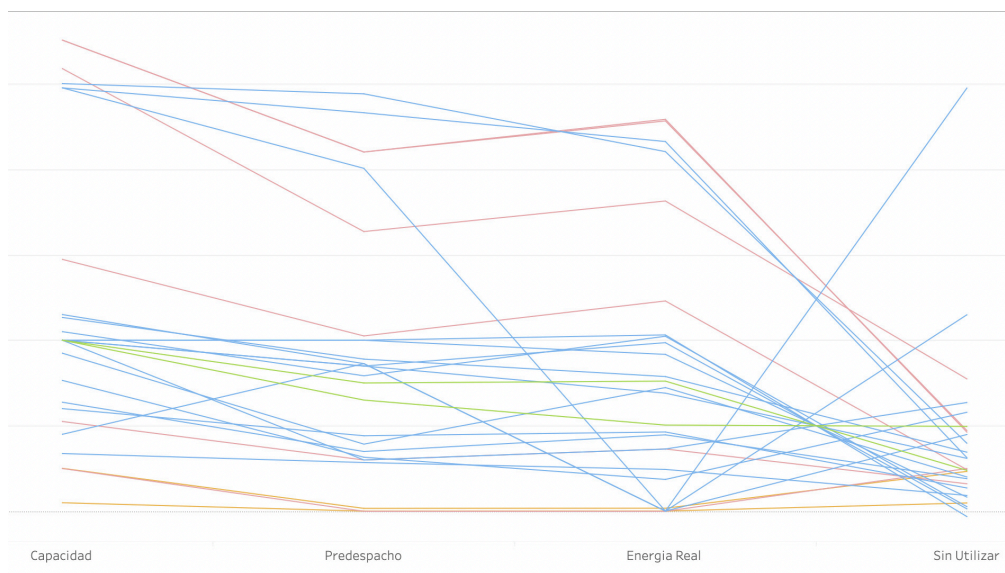


Figura 42. Parallel Coordinates: relación entre las variables.

Stacked Area Graph

Permite compara y observar las diferencias y patrones presentes en las variables de predespacho y energía real de un grupo de centrales seleccionadas en un periodo de tiempo establecido. La figura 42 muestra esta variables en un grupo de centrales eólicas, geotérmicas e hidroeléctricas.

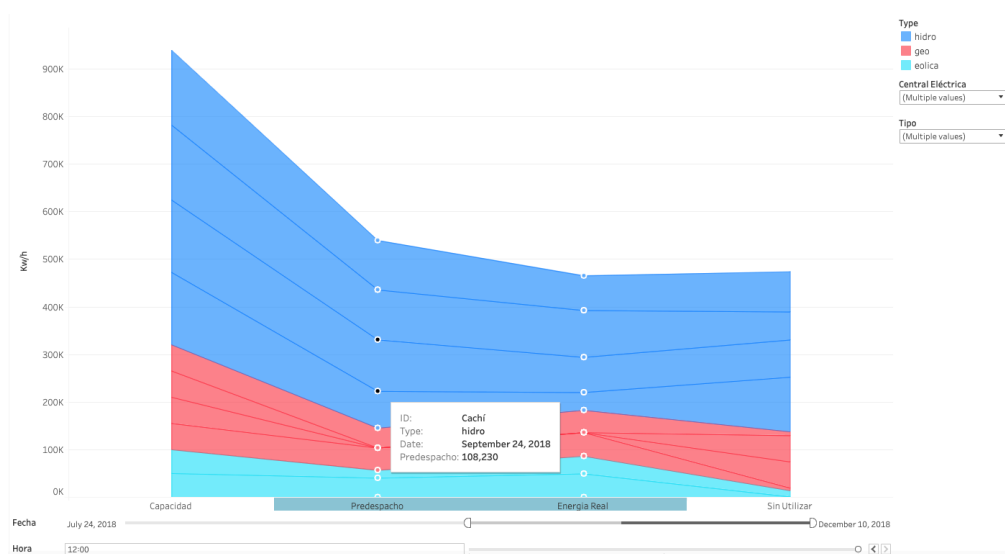


Figura 43. Stacked Area Graph: relación entre las variables.

12.2.4 Época del año con mayor producción por cada central

Stream graph

Permite visualizar el comportamiento de producción energética de una o varias centrales eléctricas en el tiempo. Por ejemplo en la figura 44 , se muestra la producción de energía para un día en específico de todas las centrales eólicas del país.

Esta visualización no permite visualizar un conjunto de días al mismo tiempo, por lo que la interpretación podría llevar a errores en caso de querer analizar grandes periodos de tiempo.

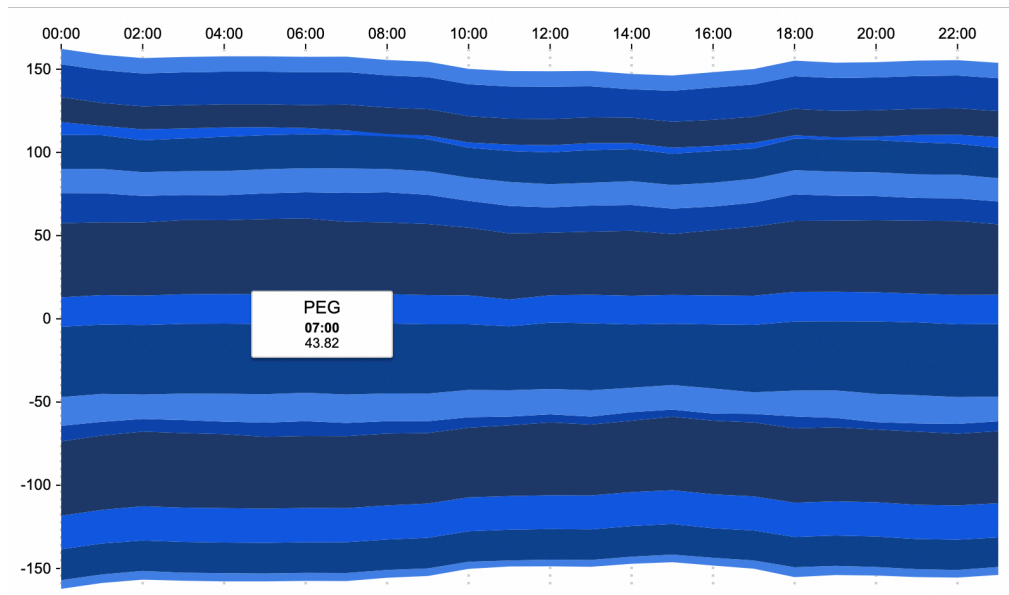


Figura 44. Stream Graph: producción de un día específico de las centrales eólicas.

Parallel Coordinates

Con esta visualización se puede observar el comportamiento de producción de las centrales en un periodo de tiempo, por ejemplo un día y hora específica o un grupo definido de días.

En la figuras 45, se analizan la producción de todas las centrales eólicas los días 23 de julio y 10 diciembre, obteniendo como resultado el día con un grado de productividad mayor.

Esta visualización también permite visualizar meses o semanas.

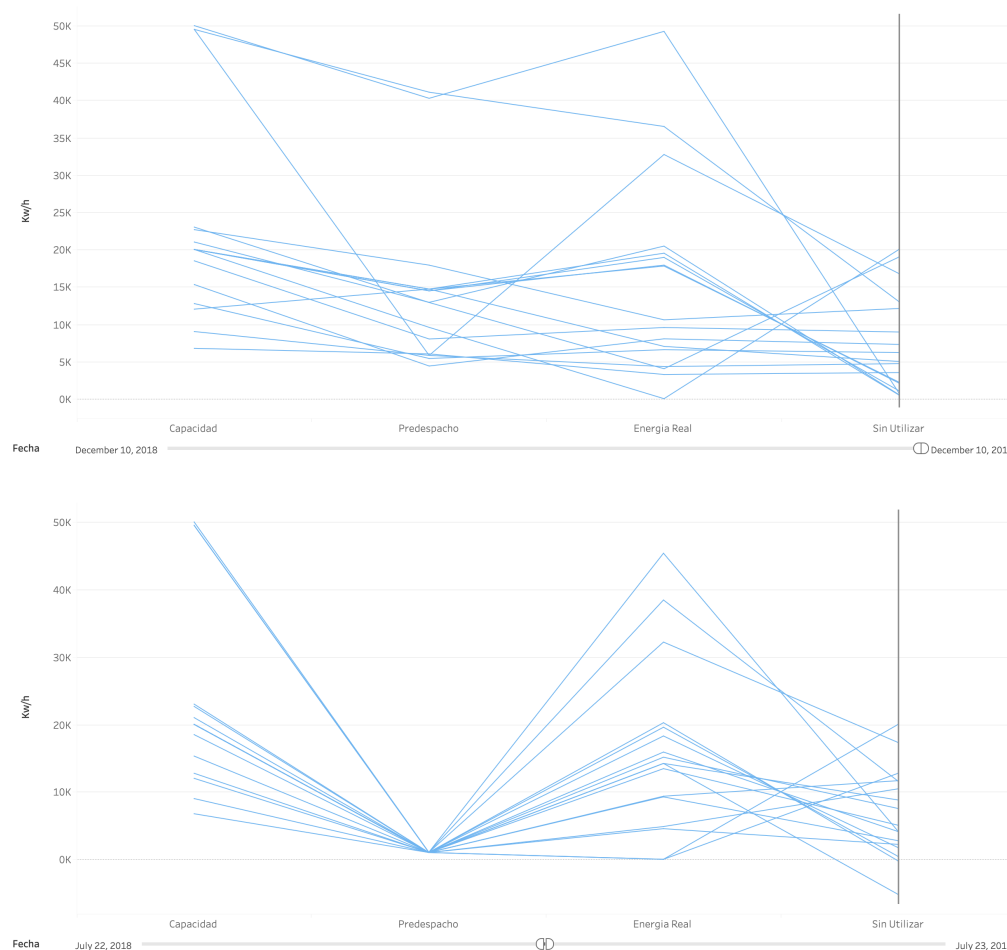


Figura 45. Parallel Coordinates: producción eólica en los días 23 de julio y 10 de diciembre.

Stacked Area Graph

De igual forma que que los parallel coordinates, este tipo de visualización permite observar el comportamiento de producción en periodos de tiempo especificados por el usuario.

En la figura 46 se puede observar la energía real generada el día 22 de julio del año 2018 de diferentes centrales.

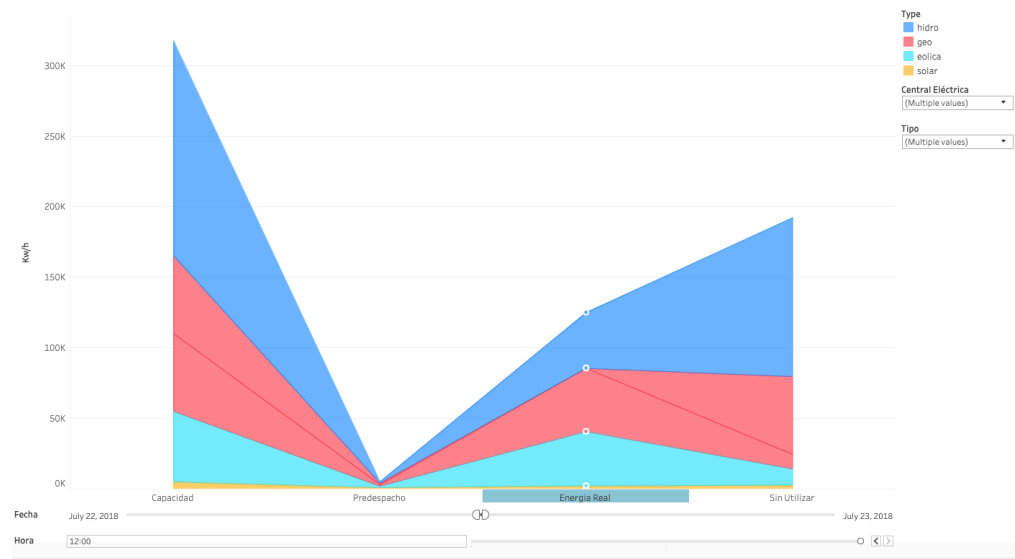


Figura 46. Stacked Area Graph: energía real generada el día 22 de julio 2018..

12.3 Matriz de Selección del Paradigma

Con la finalidad de evaluar los paradigmas seleccionados estos se sometieron a dos evaluaciones según lo estudiado y propuesto por varios autores (Hernandez-Castro et al., 2016), la primera correspondiendo a los casos de uso y la segunda a requerimientos de la visualización.

12.3.1 Matriz de evaluación según casos de estudio

Para el análisis de cada uno de los paradigmas con respecto a los casos de uso, se someten a una evaluación cuantitativa, tomando como base el poder comprender la información, facilidad de comparación entre las variables y la cantidad de niveles de navegación para el usuario. Se utiliza una escala de 1 a 5, siendo 5 la mayor puntuación)

Caso de Uso	Stream Graph	Stacked Area Graph	Parallel Coordinates
Caso 1: Visualizar cada central eléctrica	4	3	5
Caso 2: Observar el grado de subutilización de las centrales	4	2	5
Caso 3: Relación entre el predespacho y la energía real	4	3	5
Caso 4: Época del año con mayor producción por tipo de central	5	3	4
Total	17	11	19

Tabla 4. Matriz de evaluación según casos de uso.

12.3.1 Matriz de evaluación según requerimientos de visualización

Cada uno de los paradigmas es sometido a una evaluación de criterios que debería cumplir la visualización, de modo que el uso por parte de los usuarios sea óptimo. Estos criterios corresponden a facilidad de lectura, uso intuitivo de la herramienta, escalabilidad de la información, realizar comparaciones de forma intuitiva, así como visualizar información global y específica. Se utiliza una escala de 1 a 5, siendo 5 la mayor puntuación).

Requerimiento	Stream Graph	Stacked Area Graph	Parallel Coordinates
Facilidad de lectura	2	2	4
Escalable	3	2	4
Uso intuitivo	4	2	4
Permite comparaciones	4	2	5
Visualización Global	5	4	4
Visualización Específica	5	3	5
Total	23	15	26

Tabla 5 . Matriz de evaluación según requerimientos de visualización.

En las anteriores tablas de evaluación, el paradigma Parallel Coordinates obtuvo la mejor puntuación y es seleccionado para utilizarlo en la implementación final de la visualización.

13. Implementación

La implementación final de la visualización se lleva a cabo utilizando Tableau Public. Los datos son utilizados en formato .CSV lo cual facilita el manejo de gran cantidad de estos.

Debido a la gran granularidad de los datos, se decide trabajar con datos distribuidos en el año 2018, para un total aproximado de 35 000 datos.

Además, una vez decidido el paradigma de visualización se decide realizar una visualización secundaria derivada de los parallel coordinates, apoyando en la visualización de la generación real a través del tiempo por cada central obteniendo patrones y tendencias de los datos.

13.1 Definición de aspectos visuales

La estética visual es definida por medio de color y líneas:

Cromática

La cromática presente en la visualización es definida según el tipo de fuente utilizada para generar la energía, como se muestra en la figura 47.

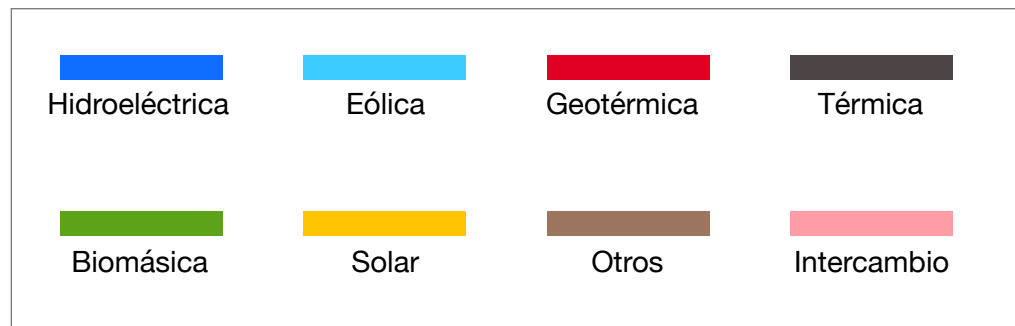


Figura 47. Parallel Coordinates.

El resto de la información y elementos se disponen en escala de grises, de modo que no interfieran con la visualización de los datos por medio de líneas en el parallel coordinates.

Líneas

Las líneas utilizadas para unir las variables en el parallel coordinates se disponen en un grosor de 2 pixeles de forma que estas puedan ser observables en casos donde se muestra gran cantidad de información.

13.2 Visualización de Datos

A continuación se muestra la visualización de datos final, la cual muestra en su eje vertical un rango de kW/h, en su eje horizontal las variables de capacidad instalada, predespacho, generación real y energía no generada (capacidad sin utilizar).

El periodo de horas o días a visualizar es definido por medio de líneas de tiempo interactivas que permiten definir un solo día/hora o un rango de estos.

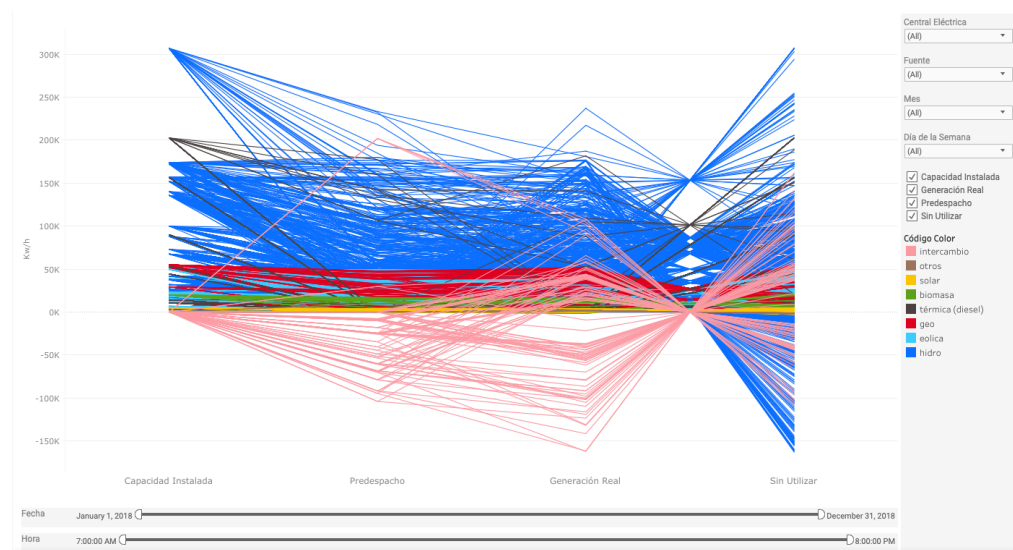


Figura 49 . Implementación final, en la cual se visualizan los datos de energía del año 2018.

Una serie de filtros pueden ser activados para obtener información en detalle ocultando y mostrando grupos de líneas, de forma que el usuario puede decidir la central eléctrica, el tipo de fuente, el mes y el día de la semana que desea visualizar y comparar.

En la figura 50 se observa cómo los parallel coordinates permiten realizar **“brushing”** una selección de las líneas que se desean visualizar, facilitando visualizar el recorrido de estas en casos de gran cantidad de información.

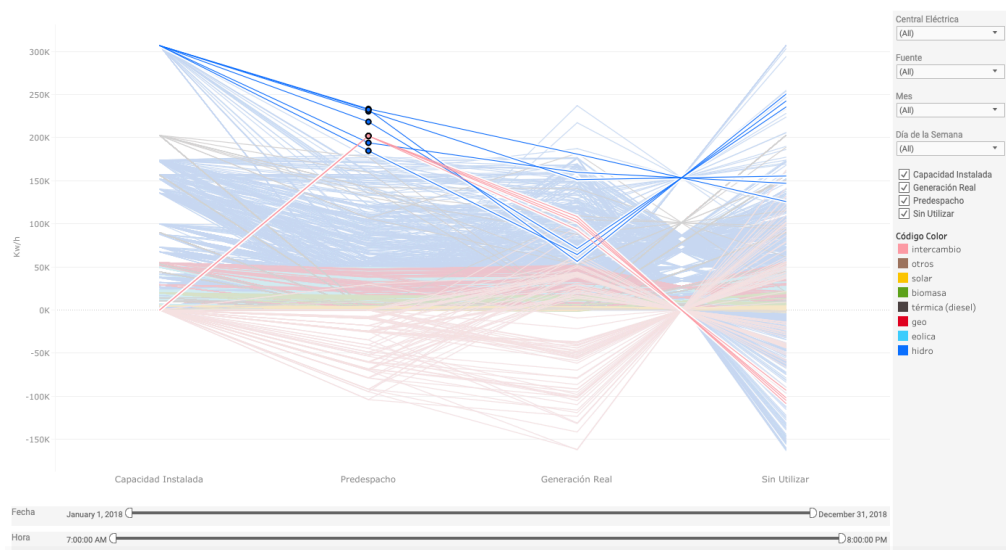


Figura 50 . Técnica de brushing en los parallel coordinates.

Otra de las técnicas para disminuir la cantidad de información es seleccionar los recorridos (líneas) que se desean visualizar y excluir el resto de información, esto se muestra en la figura 51.

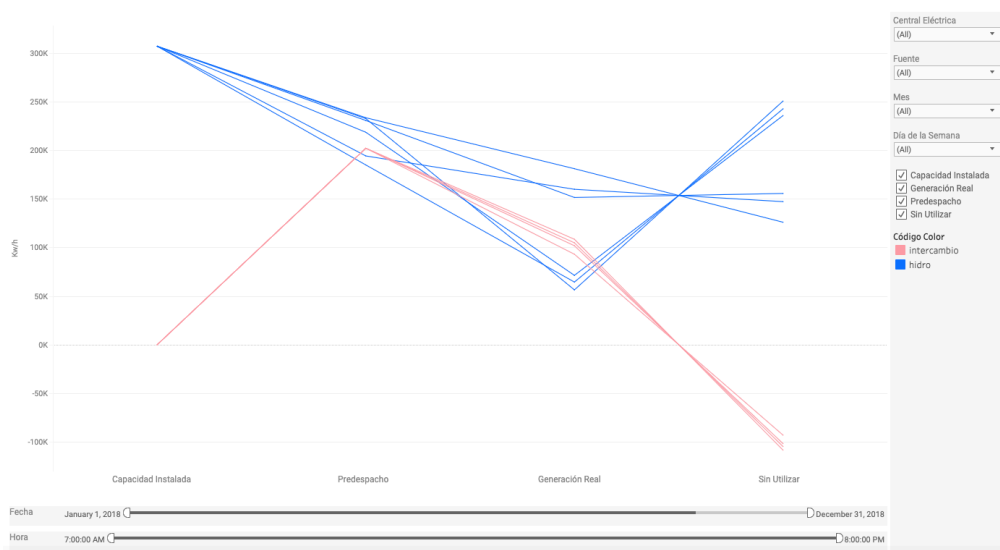
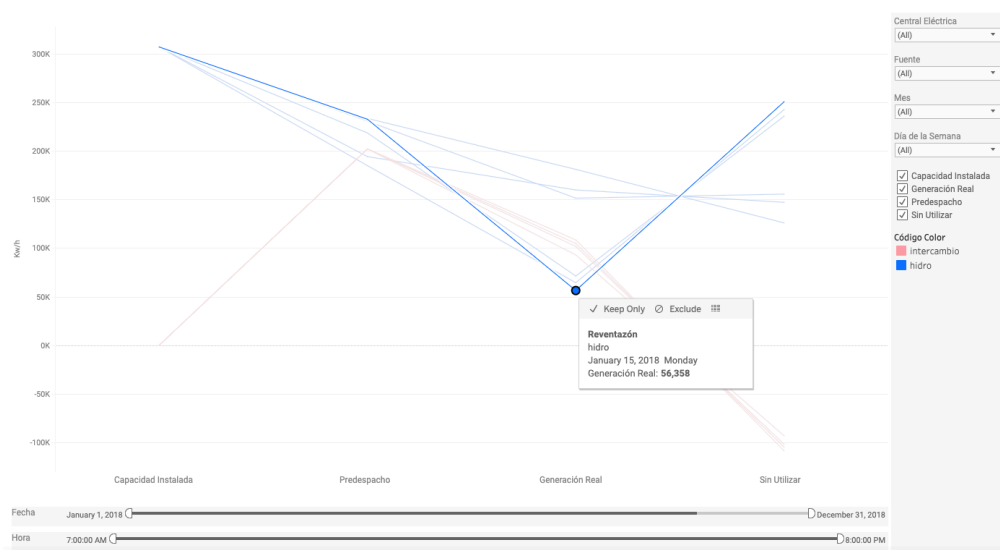


Figura 51. Exclusión de información que no se desea visualizar.

El usuario puede visualizar información específica de cada uno de los recorridos por medio de tooltips que se muestran al



realizar “mouseover” y “mouseclick”.

La visualización secundaria realizada muestra en su eje vertical un rango de kW/h y en su eje horizontal los días del año, permitiendo observar el comportamiento de cada una de las centrales en periodos de tiempos definidos.

En este caso cada una de las coordenadas o líneas corresponde a una central y responden a la totalidad de energía generada por cada una en un día.

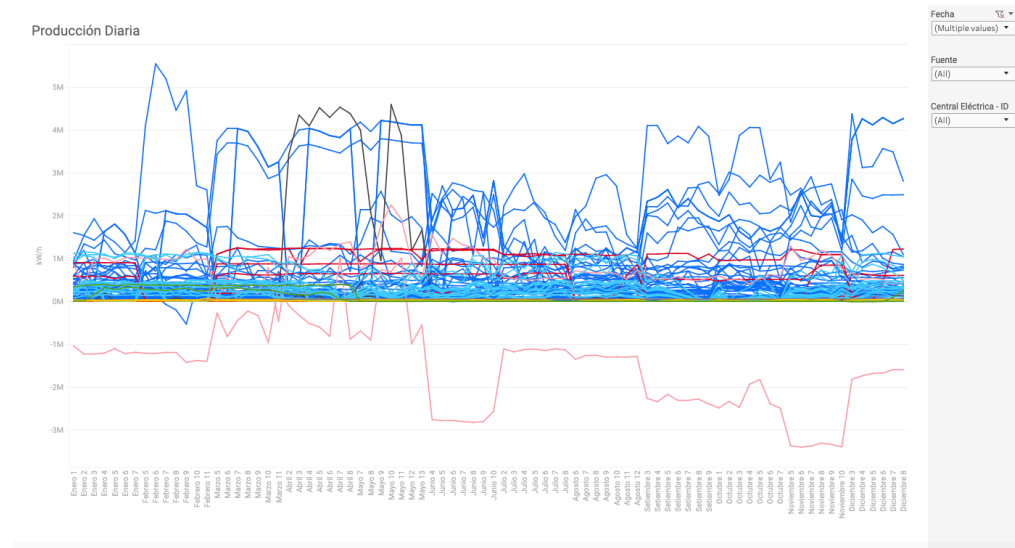


Figura 53. Visualización secundaria derivada de los Parallel Coordinates

Esta visualización permite filtrar información según día del año (fecha), fuente utilizada y el ID de la central, por ejemplo, en la figura 54 se puede visualizar el comportamiento de las centrales eólicas en diferentes días del primer semestre del año 2018. Además información en detalle puede ser observada por medio de tooltips al hacer “mouseover” o “mouseclick” sobre las líneas.

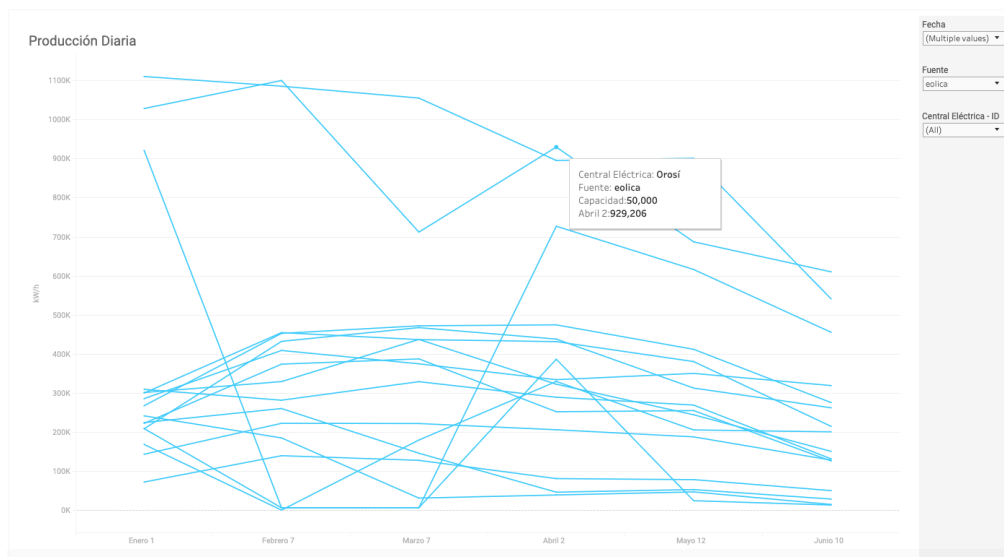


Figura 54 . Comportamiento de las centrales eólicas durante días definidos del primer semestre del año.

14. Validación y resultados.

La visualización final muestra información de 35 mil datos energéticos del año 2018, sin embargo la herramienta es escalable a cualquier periodo de tiempo con los datos disponibles.

Para validar la propuesta final, se definieron casos de uso con la finalidad de evaluar las diferentes opciones de visualizar los datos.

14.1 Casos de estudio

14.1.1 Determinar cual central tiene el grado mas alto de sub utilización

En este caso basta con observar cuál coordenada se encuentra en el punto más alto de la columna de Capacidad Sin Utilizar.

14.1.2 Determinar cual central produjo mas energía

De igual forma que el caso anterior la visualización es clara, por lo que solamente se debe visualizar cual coordenada se encuentra en el punto mas alto de la columna de Generación Real.

14.1.3 Buscar una central eléctrica por su nombre

La visualización presenta una barra de búsqueda en cada uno de sus filtros, en los cuales se puede escribir la información deseada.

14.1.4 Conocer la fecha y hora específica de las coordenadas

Para ver la fecha y hora específica, así como otra información de una de las coordenadas, se puede realizar mouseover o mouseclick. Para visualizar múltiples coordenadas se pueden seleccionar or medio de la técnica de “brushing” y luego hacer mouseover sobre cada una de ellas.

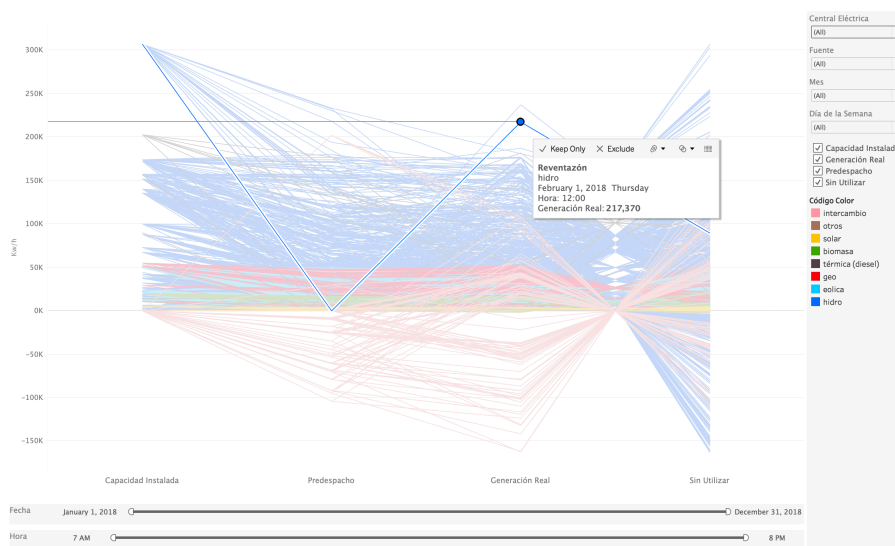


Figura 55 . Tooltip que indica fecha y hora específica de las coordenadas.

14.1.5 Comparar dos centrales en un periodo de tiempo específico.

Los usuarios deben seleccionar las centrales que desean comparar e indicar el periodo de tiempo que visualizar. En el caso de la figura 56, se comparan las centrales hidroeléctricas de Arenal y Angostura el 15 de noviembre a las 7 am.

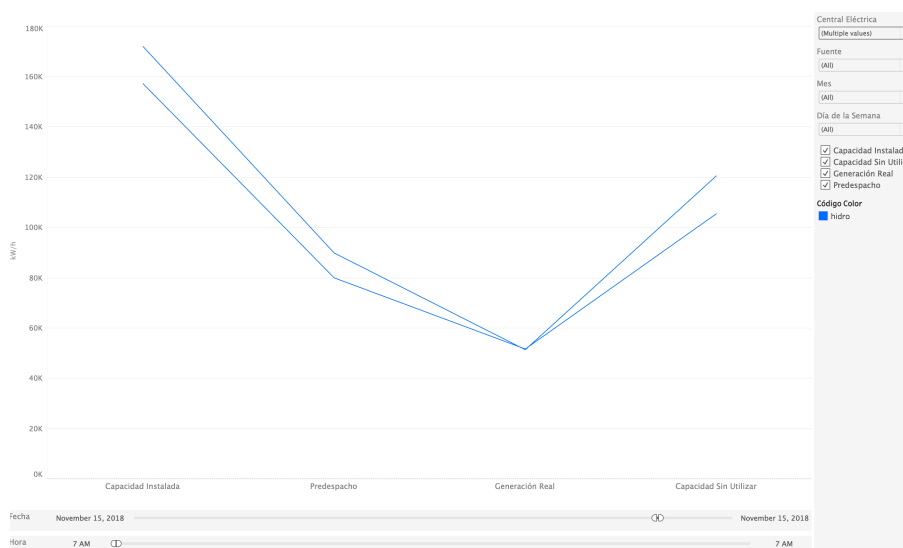


Figura 56 . Comparación de las centrales Arenal y Angostura.

14.1.6 Comparar dos variables por separado

La visualización permite ocultar las columnas (paralelas) de forma que basta con ocultar aquellas que el usuario no desea ver. Esto permite disminuir el ruido en la composición.

En el caso de la figura 57, se ocultaron las variables de capacidad instalada y capacidad sin utilizar.

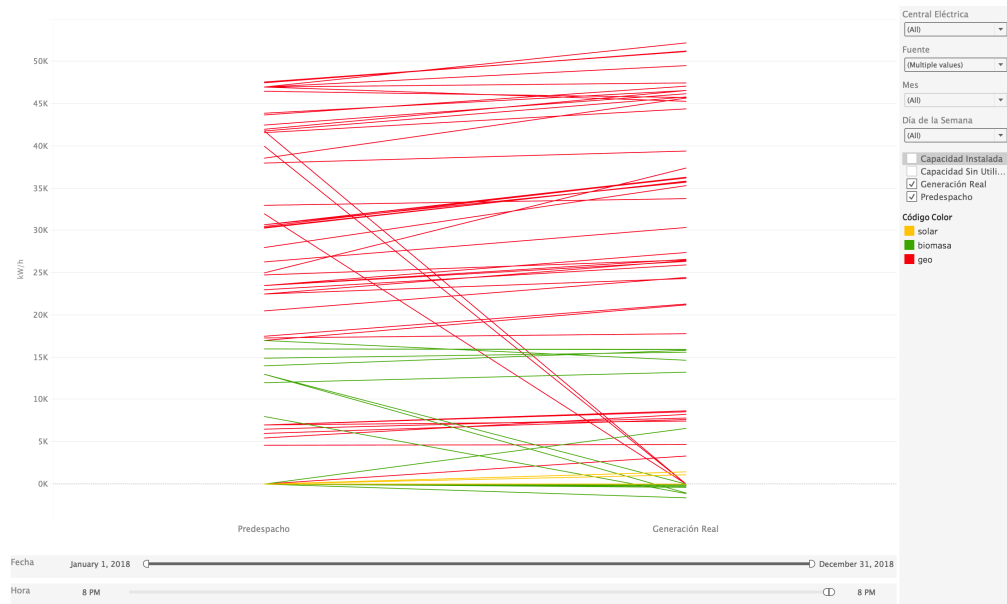


Figura 57. Comparación de las variables de Predespacho y Generación Real.

14.1.7 Visualizar el mismo día de la semana a través un mes a una hora específica

El usuario debe elegir el día de la semana, el mes y determinar la hora que desea visualizar, por ejemplo en la figura 58, se visualizan los lunes de enero a las 7 am de las diferentes centrales eólicas.

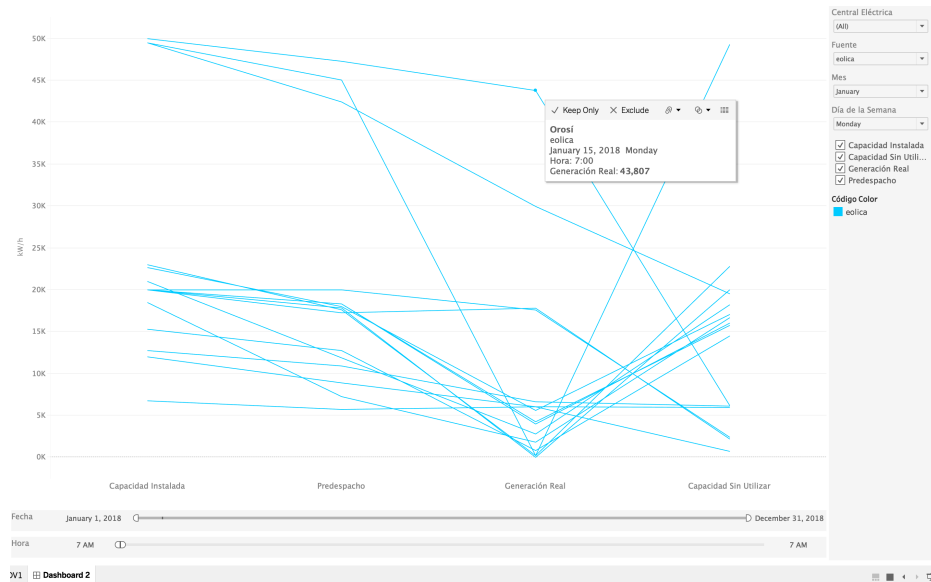


Figura 58 . Lunes del mes de enero de las centrales eólicas.

14.1.8 Visualizar el comportamiento de un tipo de fuente en un periodo específico de tiempo.

Para esto el usuario puede dirigirse a la visualización secundaria, donde puede determinar los días que desea visualizar, al igual que el tipo de fuente.

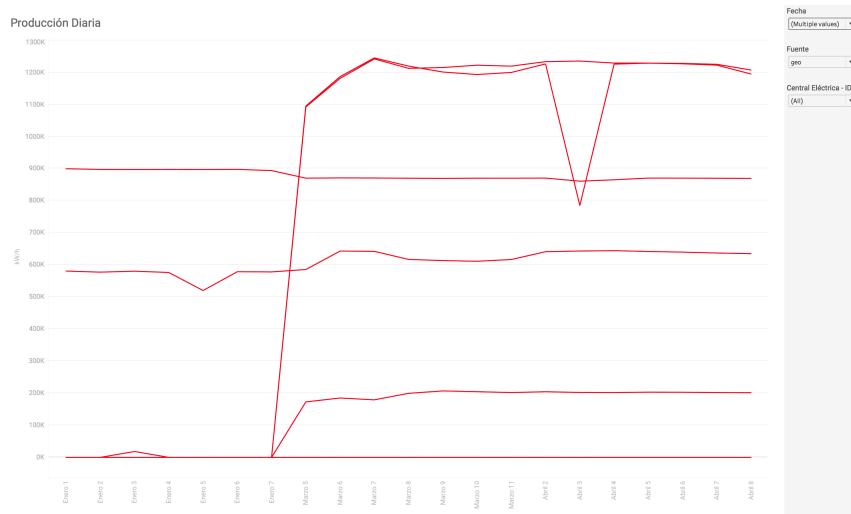


Figura 59 . Comportamiento de la producción de energía geotérmica durante diferentes días del año.

14.1.9 Visualizar en cual época del año se produce mas energía para un tipo de fuente en específico

El usuario debe dirigirse a la visualización secundaria, elegir el tipo de fuente que desea visualizar y observar a qué día corresponden los picos mas altos de generación de energía.

En el caso de la figura 60, se observa que los primeros 4 meses del año son los de mayor producción para las centrales que utilizan biomasa como fuente.

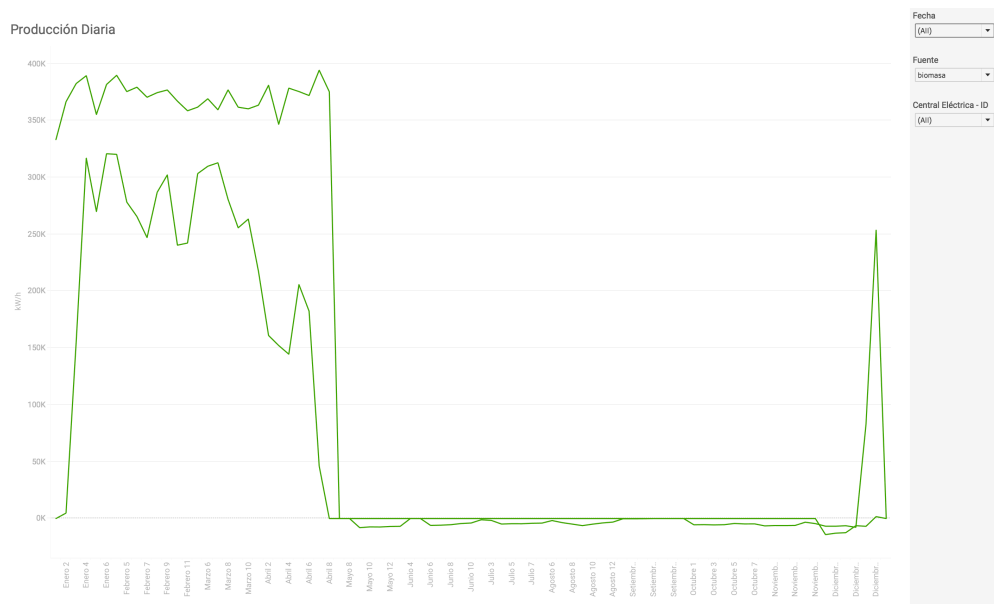


Figura 60. Producción de energía a partir de biomasa durante el año 2018.

14. Hallazgos

Dentro de los hallazgos principales se encuentra que el uso de fuentes no renovables como la térmica se mantiene en números muy bajos durante el año, excepto en los meses de marzo y abril donde la producción de energía hidroeléctrica disminuye.

Se confirma que la producción de energía eólica es baja durante los meses de mayor precipitaciones.

La energía a partir de biomasa es aprovechable solamente durante los meses de cosecha, manteniéndose inactiva durante los meses de siembra.

La energía geotérmica es una de las más estables en producción durante todo el año, por lo que debería considerarse como alternativa para disminuir el uso de fuentes no renovables en los meses más secos del año.

Por último, es observable que constantemente aportamos energía a través de la frontera norte, y generalmente recibimos energía de la frontera sur, siendo un patrón repetitivo a través del tiempo.

15. Conclusiones

La visualización propuesta permite incrementar el uso de los datos y dar mas valor a estos, permitiendo comprender de una mejor forma la matriz eléctrica costarricense.

Se determina que la producción eléctrica se ve afectada por factores y variables externos como el clima.

El uso de líneas a través de diferentes puntos permite tener una perspectiva de flujo de los datos según las variables visualizadas.

La herramienta de visualización permite responder la pregunta objetivo principal “¿Cuál es el grado de sub utilización de las centrales productoras de energía?”

Esta visualización permite visualizar datos de energía históricos, apoyando en la toma de decisiones por parte del dominio experto.

16. Trabajo a futuro

Se propone migrar los datos CSV a SQL, de forma que se puedan realizar mas relaciones entre datos, así como manejar mayor cantidad de estos.

Alimentar la visualización con datos de otros años, permitiendo conocer el cambio en el uso de energía renovables en Costa Rica.

Aumentar la capacidad de la visualización secundaria, de modo que se pueda elegir observar otras variables como el predespacho técnico nacional.

Relacionar los datos utilizados con otros recolectados por el CENCE, por ejemplo datos de mantenimientos programados.

17. Referencias

Akhavan-Hejazi, H., & Mohsenian-Rad, H. (2018). Power systems big data analytics: An assessment of paradigm shift barriers and prospects. *Energy Reports*, 4, 91–100

BBC. (2017) Cómo hizo Costa Rica para pasar 250 días sin utilizar combustibles fósiles para su electricidad. Recuperado el 3 Setiembre del 2018, de [ht-tps://www.bbc.com/mundo/noticias-38508608](https://www.bbc.com/mundo/noticias-38508608)

BBC. (2015) Cómo hace Costa Rica para producir toda su electricidad de manera limpia. Recuperado el 3 Setiembre del 2018, de <https://www.bbc.com/-mundo/noticias-38508608>

Centro Nacional de Control de Energías. (s.f.-b). Generación Real [Conjunto de datos]. Recuperado 1 febrero, 2019, de <https://apps.grupoice.com/CenceWeb/CencePosdespachoNacional.jsf?init=true>

Centro Nacional de Control de Energías. (s.f.). Predespacho Técnico Nacional [Conjunto de datos]. Recuperado 1 febrero, 2019, de <https://apps.grupoice.com/CenceWeb/CencePredespachoTecnicoNacional.jsf>

Cottafava, D. Sonnetti, G. Gambino, P. Tartaglino, A. (2018). *Explorative Multidimensional Analysis for Energy Efficiency: DataViz versus Clustering Algorithms*. *Energies: Open Access Journal*. May 2018. Suiza.

CRHoy.com. (2017). Costa Rica acumula 300 días con electricidad 100% renovable. Recuperado el 10 de Noviembre del 2018, de <https://www.crhoy.com/am-biente/costa-rica-acumula-300-dias-con-electricidad-100-renovable/>

Dirección Comunicación e Identidad Corporativa – ICE. (2015). *Costa Rica: Matriz Eléctrica, un modelo sostenible único en el mundo*. Recuperado de https://www.grupoice.com/wps/wcm/connect/8823524c-7cc7-4cef-abde-a1f06e14da0e/matriz_folleto_web2.pdf?MOD=AJPERES&CVID=I8SK4gG

Escobedo, G., Jacome, N., & Arroyo-Figueroa, G. (2017). Big Data & Analytics to Support the Renewable Energy Integration of Smart Grids Case Study: Power Solar Generation. *Science and Technology Publications*

Gobierno CR. (2017). Generación renovable de 2017 es la más alta de los últimos 30 años. Recuperado el 10 de octubre del 2018, de <http://gobierno.cr/generacion-renovable-de-2017-es-la-mas-alta-de-los-ultimos-30-anos/>

Gobierno CR. (2017) Instalación eólica creció 545% en última década. Recuperado el 10 de octubre del 2018, de <http://gobierno.cr/instalacion-eolica-crecio-545-en-ultima-decada/>

Gobierno CR. (2017) Costa Rica tiene la matriz eléctrica más limpia del planeta. Recuperado el 12 de octubre del 2018, de <http://gobierno.cr/costa-rica-tiene-la-matriz-electrica-mas-limpia-del-planeta/>

Hernandez-Castro, F., & Monge-Fallas, J. (2016). What for: classification of visual paradigms. *PONTE*, 72(7), 56–60.

Hernandez-Castro, F., Monge-Fallas, J., Gonzalez Villalobos, S., Barquero Rodriguez, E., & Esquivel Piedra, J. (2016). Traffic Data Visualization in Costa Rica: A Visualization of top 100 routes with the highest traffic density in Costa Rica. *PONTE*, 72(11), 263–272.

Klump, R. P, & Weber, J. D. (2002). Real-Time Data Retrieval and New Visualization Techniques for the Energy Industry. *Proceedings of the 35th Hawaii International Conference on System Sciences*, , 1–6.

Ribbecca, S. (s.f.). Showing Data Over Time: The Data Visualisation Catalogue. Recuperado 10 marzo, 2019, de <https://datavizcatalogue.com/search/time.html>

Teletica.com. (2017). Costa Rica tiene 326 días de generar electricidad mediante energía limpia. Recuperado el 10 de Noviembre del 2018, de https://www.teletica.com/180994_costa-rica-tiene-326-dias-de-generar-electricidad-mediante-energia-limpia

Quesada, J. (2014) Generación eléctrica en Costa Rica. ARESEP Autoridad Reguladora de Servicios y Universidad de Costa Rica. San José, Costa Rica.

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, 215–225.